SCHOOL OF OPERATIONS RESEARCH
AND INFORMATION ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853

Financial Engineering Project

ALPHA CLONING - FOLLOWING 13F FILLINGS

Presented to the Faculty of the Graduate School
of Cornell University
in partial fulfillment of the requirements for the
Master of Engineering Degree

Bowen Gao, bg453
Ruyu Kong, rk742
Yuhan Li, yl2794
Duo Sun, ds2324
Haoran, Wang, hw629
Ziming Zeng, zz494

December 2018

Faculty Advisor's Signature

Date

Cornell University

ORIE 5220 Rebellion Research Group Project

Alpha Cloning - Following 13F Filings

Authors:

Bowen Gao bg453

Ruyu Kong rk742

Yuhan Li yl2794

Duo Sun ds2324

Haoran Wang hw629

Ziming Zeng zz494

Sponsors:

Alexander Fleiss

Han Cui

Academic Advisor:

Prof. Sasha Stoikov

December 8, 2018

Contents

1 Background

SEC Form 13F is a quarterly report that is filed by institutional investment managers with at least $100 million [1] in equity assets under management, discloses their U.S. equity holdings to the Securities and Exchange Commission (SEC) and provides insights into what the smart money is doing. Firms that are required to file 13F include mutual funds, hedge funds, trust companies, pension funds, insurance companies and registered investment advisers [2].

SEC Form 13F filings provide investors with an inside look at the holdings of Wall Street's top stock pickers and their asset allocation strategies. Individual investors also scrutinize 13F filings to generate investment ideas. By looking at the top holdings of some managers, investors hope to put together a best-ideas portfolio without paying management fees. Hence, part of our project goal is to select the top performance funds and selectively Copycat their portfolio by analyzing the past quarters return and money inflow.

However, the potential risk behind is the tendency of hedge funds to borrow investment ideas from each other, and the bandwagon effect can lead to crowded trades and overvalued stocks. Hence, the other emphasis of our project is to use Machine Learning techniques together with the 13f data and price data to predict stock return and build portfolio.

In this paper, we will first give details of how we warehouse and analyze the 13f data and price data, then we will zoom in both strategies and build our portfolio. Lastly, we will compare the two strategies in a backtester.

[1] From https://www.sec.gov/pdf/form13f.pdf

[2] From https://www.investopedia.com/terms/f/form-13f.asp

2 Data Cleaning

2.1 Stock Price

We used stock prices provided by Rebellion Research. It includes fundamental stock level data from 2001/01/02 to 2018/07/30. There are in total 39879 distinct stocks and on average 19994 distinct stocks for each quarter. The price data has the following useful columns:

- pSP_CUSIP: stock CUSIPs in character with length 9 and were used to merge with 13f table
- pSP_TICKER: stock tickers with length 6 and were used to find industry related ETF used in Classification section
- pSP_DATE: trading dates in datetime format
- pSP_VOLUME: stock trading volume. We used this value during the final stage of building the portfolio that since we did not want the portfolio to include too many stocks, we limited the number of stocks to 200 and selected those with largest trading volume as stocks with small trading volume tend to be volatile.
- pSP_CLOSE: adjusted closed price, the main column we used in our modeling process to calculate fund return and portfolio return. Note here that although in the 13f fund data, we have market value data available for each filling that we can aggregate to get fund market value, we did not use that data because they contained many errors as will be explained in next subsection.

Although the price data is relatively clean, there are still some missing values issues in adjusted closed price and volume. We first extracted all the distinct dates, sorted adjusted closed price and volume by ascending date order so that we can see when values are missing.

Then, to deal with missing values, there are three general solutions. First, we could look up from other data source and try to find the missing value. We applied this method to fill out missing ETF data. Since we used 10 ETF data to perform classification which was the foundation in our stock-level predicting model, it is important to have correct ETF price data. Specifically, in the original dataset, most ETF we used had missing values since 2017/05/12, so we used Yahoo Finance data and merged with original price data. Second, to deal with adjusted closed prices, we filled the missing valued with the adjusted closed price from previous day. We could have used interpolation or mean value of previous several days but due to the size of the data, we simply used previous day data. Third, to deal with volume, if on the same day volume and adjusted closed price are missing at the same time, it is very likely due to a data issue then we fill the volume with the previous day volume. If on the same day adjusted closed price is not missing but volume is, it is likely that there was no trading on that day and we filled the volume by 0.

2.2 SEC Form 13f

Since 13 filings is a quarterly reported file, our data cleaning process is consistent with their quarterly period. Same as other data cleaning process, we first dropped the duplicates and missing values. Then we looked at abnormal values, for this step we dropped records which have zero market value and zero quantity. After the general data cleaning, we further cleaned data based on both personal and regulatory requirements.

For each quarter, We chose funds which had the number of stock holdings between 20 and 200. We believed it is a reasonable range since holdings greater than 200 is too diverse and maybe institutional managers just followed the majority of market thus it cannot showed their stock picking skills or preference. We also eliminated funds with their stock holdings less than 20 since these funds are too concentrated and they can't not provide enough diversification portfolios.

In addition, according to SEC requirements, institutional investment managers with at least 100 million in equity should disclosed their U.S equity holdings. Another major requirement is that these filings should be filed within 45 days of the end of a calendar quarter, or if that day falls on holiday or weekends.[3] Referenced by this two major requirements, we kept the funds that had their market value between 100 and 500 millions. And the time difference between filing date and period end date should be less than 47 days.

Finally, we also required that funds should have exactly 1 filing point for each quarter. SEC allows institutional managers to revise and amend reported positions. For our analysis, we only considered the funds without filing emendation. Also this step, we implicitly deleted the funds that have missing reported filings. This additional requirement benefit for our return analysis.

After quarterly data cleaning and analysis,we can get some insights for distinct funds remained for each quarter. We found that there are only below 30 funds remained before 2013/06/30. Based on the reference from SEC, we knew that on 2013/05/20, the text-based ASCII format for 13F filings was discontinued and replaced with an online form. [4] Due to the incompletely data before 2013 May, we decided to only contain the period June 30 2013 to Dec 31 2017.

In order to consider the persistence of the funds, we decided to keep the funds who consistently meet the requirement for at least 1 year, then we can get the union of these funds. We defined these universe of funds as survival funds and used for our further analysis. With the initial 8100 distinct funds, after all the data cleaning requirements, we ended up with

[3]From https://https://www.sec.gov/divisions/investment/13ffaq.htm

[4]Fromhttps://www.lexology.com/library/detail.aspx?g=71f232e8-5c96-4c4e-b9d2-99bba6089232

1000 distinct funds.

The following table showed our typically quarterly data cleaning steps and statistics of funds remained after the cleaning.

Table 1: Data Cleaning Criteria and Results(2014-06-30)

Steps	Criteria	Distinct Funds	Data Remaining
1	Drop zeros and duplicates	4276	99.57%
2	Drop zero Market cap and iQTY	4220	98.69%
3	Drop abnormal stock price	4220	98.69%
4	Drop extra filling dates for each period	4215	98.57%
5	Keep funds holding 20-200 stocks	2348	54.91%
6	Keep funds market value 100M-500M	1331	31.12%

3 Data Analysis

3.1 Classify by Industry Sector

The most straight-forward way to classify funds is to put them into different sectors. However there is no direct way to look up each fund industry since majority of them is very diverse in their investment. A very intuitive approach is to investigate the industry of their invested stocks. We used Bloomberg API to find the sector information for each stock. After we knew the sector of each stock belongs to, it is straight-forward to group the stocks of each fund according to their sectors. This idea is so-called subfunds. For example, after we know fund A's holdings, we group all the technology stocks as a subfund, group all the financial stocks as a subfund, and so on and so forth.

However, not all the stocks' sector can be found from Bloomberg API. This will cause the sub-funds to be biased. Hence, we switched to other two consistent methods. The main idea is to run correlation of funds or stocks with ETFs and then classify the stocks or funds into different sectors with the largest correlation. We first tried to run the correlation between funds return and each 10 EFTs return from 2013/06/30 to 2017/12/31. For all the 10 industries, we assigned the fund to the industry which can give us the highest correlation. However, this may still cause bias, since it is possible that fund's return indeed has good correlation with some specific sector but it may not heavily hold them. Actually, the combination of the returns of the funds' majority holdings shows a great correlation to a particular sector. Also the majority of the funds have very diverse holding, it is unfair to assign a whole fund into a sector. Another important reason is that it is required to have at least 3 data points to run the correlation, which does not applicable to some funds.

Hence, we combined the idea of sub-funds and correlation. We run the correlation of each stock return with the 10 ETFs return, then classify each stock to the particular ETF which gives the highest correlation. Then we split each fund into sub-funds, with each sub-funds contain stocks all in specific industry sector.

Below is the 10 ETF sectors that we chose.

Table 2: EFT Industry

ETFs	Class
XLF	Financial Select Sector SPDR Fund
XLY	Consumer Discretionary Select Sector SPDR Fund
XLK	Technology Select Sector SPDR Fund
XLB	Materials Select Sector SPDR Fund
XLI	Industrial Select Sector SPDR Fund
XLE	Energy Select Sector SPDR Fund
XLU	Utilities Select Sector SPDR Fund
XLV	Health Care Select Sector SPDR Fund
XLP	Consumer Staples Select Sector SPDR Fund
VOX	Communications Equities

3.2 Classify by Value and Size

Without a doubt, size and value are other crucial indicators for a fund. Value stocks[5] are less expensive than the broader market. One reason for this is that value companies may be riskier because they have more leverage than the broader market, visible in their higher debt-to-equity ratios. Hence funds invest heavily in value stocks tend to have relatively low profit margins, implying that they are not as flexible and profitable as the broader market, and have lower expected earnings growth. Big size stocks often regard to bigger market capitalization in a particular area. Those stocks often have steady return and less volatile. Funds hold more stocks in big size imply they are very conservative in investment and have relatively stable return.

The idea of this classification is more or less similar to the industry classification. Here we instead using the Size and Value ETFs to do the correlation for each stock and split funds into sub-funds.

Below is the 10 ETF sectors that we chose.

[5]Fromhttps://www.oppenheimerfunds.com/advisors/article/when-evaluating-factors-consider-seasonality

Table 3: EFT Industry

ETF	Class
SPY	Large Cap Blend Equities
QQQ	Large Cap Growth Equities
IWD	Large Cap Value Equities
IJH	Mid Cap Blend Equities
IWP	Mid Cap Growth Equities
IWS	Mid Cap Value Equities
IWM	Small Cap Blend Equities
IWO	Small Cap Growth Equities
VBR	Small Cap Value Equities

3.3 Other Classifications

3.3.1 Performance

Consistent past performance is always a good metric to start analyzing a fund. A fund with higher and consistent past performance may indicate their ability to analyze the market supply and demand and their good acumen in both bullish and bearish market. Hence, it will be useful to classify and identify such funds.

For each fund from 2013/06/30 to 2017/12/31, we recorded their return between their every two filing dates. Then we calculated the mean of their return throughout this period. Now we have the mean return of each fund, M_i in the whole period, then we will calculate the mean of their return M, which is the mean performance of all of the funds. If a single fund's mean return is higher than the product of a threshold constant c and the total mean, which is $M_i > cM$, We will classify such fund as a high performance fund and vice versa.

In our project, we arbitrarily defined two cutoffs, 1.2 and 0.8. If $M_i > 1.2M$, we classify this fund as high performance fund, if $M_i < 0.8M$, we classify this fund as high performance fund, the rest is medium performance fund.

3.3.2 Volatility

Volatility is always a proxy for risk. Investors continue to invest in high-risk stock because the potential profits these securities offer over time often surpass what is available in other asset classes. A fund heavily invested in high-risky stocks are classified as high volatility fund and vice versa.

We first classified the stocks as high or low volatility. From 2013/06/30 to 2017/12/31, we calculated each stock's adjusted close price's percentage change volatility V_i. Then we calculated the mean of all the stock's volatility V. If the stock's volatility is higher than a multiplier of the mean, we will classify this stock as high volatility. In our project, we arbitrarily defined two cutoffs, 2 and 0.5. If $V_i > 2V$, we classify this fund as high volatility fund, if $V_i < 0.5V$, we classify this fund as high volatility fund, the rest is medium volatility fund.

3.3.3 Survival Time

Survival time is another important classifier. Many institutional investors wish to invest into hedge funds on a long-term basis and they seek hedge funds likely to survive a long time and to avoid liquidation, an undesirable outcome often associated with large capital losses. Survival Analysis can help investors select funds with good long-term prospects and Longevity can ease investor concerns regarding the illiquidity of hedge funds.

We defined the survival time of each fund T_i as the time difference of its very last filing date and its very first filing date. We sort the funds by the survival time and calculated the mean survival time of all funds T. we arbitrarily defined two cutoffs, 2 and 0.8. If $T_i > 2V$, we classify this fund as high survival fund, if $T_i < 0.8V$, we classify this fund as high survival fund, the rest is medium survival fund.

3.4 Idiosyncratic Risk

According to previous research by [1], hedge funds trade in high idiosyncratic risk stocks earn significantly higher abnormal returns than hedge funds trades in low idiosyncratic risk stocks. The idea behind it is that if a fund trades those high idiosyncratic risk stocks, it is very likely to process some private information about the stock. Therefore, we creates a idiosyncratic risk feature using the following 3 steps:

1. Compute idiosyncratic risks for every stock in SP500 composites. We run the Fama-French 3 Factors model for each stock in each quarter using a rolling window of 24 months. And we take $1 - R^2$ as the idiosyncratic risk measure for each stock, $StockIdio_{stock_i,quarter_j}$

2. Aggregate stock level idiosyncratic risks to get fund level idiosyncratic risks. If a fund m holds N stocks, and the weights of each stock in each quarter is $w_{stock_i,quarter_j}$ then

$$FundIdio_{fund_m,quarter_j} = \sum_{i=0}^{N} StockIdio_{stock_i,quarter_j} \times w_{stock_i,quarter_j} \tag{1}$$

3. For each stock, average the fund-level idiosyncratic risks of those funds which are currently holding the stock. Mathematically, if there are M funds holding the stock i at quarter j then:

$$IdioAverage_{stock_i,quarter_j} = \frac{\sum_{m=0}^{M} FundIdio_{fund_m,quarter_j}}{M} \tag{2}$$

3.5 Return

Institutional trades performance metrics including fund returns, flows, and turnover ratios were calculated based on the changes in quarterly snapshots of fund holdings. When calculating holding returns, we assumed that institutions change all their positions immediately after they report their holdings, and then they hold their positions until the new quarterly holdings are observed. Using these holdings, trades, and portfolio returns information we were able to derive other additional measures like asset value, turnover, and net flows. Since 13F does not require institutions to report all their positions such as short sales, our aggregated measures did not include all the assets funds hold at the end of the quarter. Thus, all measures we constructed reflect only the views on institutional long equity holdings.

In addition to the above measures, trades are derived and classified whether they are buys or sells. For each fund in every quarter end, we compared their current holdings and holdings reported in the last quarter end. In this way, we were able to derive and classify trades to know whether those purchases and sales were incremental trades, or initiating buys or terminating sells. Usually these newly opened and closed positions contains a lot of useful information to be extracted from.

Another important thing need to be taken into consideration is that institutions could file their 13F multiple times for the same quarter. The subsequent filings following the initial ones were reported in the form of amendment document. Thus, we make sure our performance metrics were not contaminated by considering the original filings only. Usually institutions do not fully disclose some of their holdings in their initial filings because this may incur free rider problems and thus can be harmful to the profitability of their portfolios. When processing data we separated the initial filings from amendments and applied our performance metrics calculation on each data set separately.

As we have mentioned above, we treated all trades as being executed on the day right after the quarter end and held until the next quarter end. In real life, however, trades can occur at any time during the quarter. This simple assumption is quite useful because estimating the actual time of execution could be hard to implement and could cause very large bias. This assumption instead not only alleviated the bias concerns but making returns calculation much easier. Another method to deal with this problem is to use the average daily price during that quarter, and this will not change the direction of individual trades because they are computed based on the changes in shares held. In our research we choose the former method because stock prices can be very volatile in one quarter and thus the average stock price in one quarter could also be very biased.

To put it more specifically, if securities existed in an institutional portfolio at the beginning of the quarter, and were found to be unlisted in the next quarter filing for the same fund, according to our previous assumption, we regarded those securities were all sold in the beginning of this quarter and they were not included in return calculation in this period. On the other hand, if securities do not exist in the portfolio reported at the beginning of the quarter but appear on the next quarter filing, we assume those newly added securities are purchased at the beginning of the quarter and are held unchanged in this quarter until this quarter end. We thus fetched the stock prices of those securities at the beginning and the end of this quarter to calculate quarterly returns for each of those newly added securities. The third part in each institutional portfolio are securities remained in both this quarter end and the next quarter end. For those securities remained in the fund portfolio, we examined their weights changes.

For those securities with increased weights in this quarter, we assume the incremental trades were all executed at quarter start and held constant for the whole quarter. Thus we used quarter end weights for those securities to calculate their returns. And the same could be applied to securities which were partially sold during this quarter. we also used quarter end weights and stock prices at the beginning and the end to calculate returns respectively.

$$SecurityReturn_{i,t} = \frac{\text{Price}_{i,t} - \text{Price}_{i,t-1}}{\text{Price}_{i,t-1}} \tag{3}$$

$$PortfolioReturn_t = \sum_{i}^{N} \omega_{i,t} \times SecurityReturn_{i,t} \tag{4}$$

In all, institutional portfolio returns are calculated as the hypothetical holdings returns of the long equity portion only of institutional portfolio which is observed at a quarterly frequency. So 13F institutional data is not good at capturing trades executed by high frequency trading institutions.

The following time line illustrated our copycat strategy period. For example, in the period from 2017-06-30 to 2017-09-30, we ranked funds' performances on their calculations on returns or inflows. This quarter is our portfolio performance calculation period. As we referenced from 13 filings, the funds are required to submit their long-only holdings no more than 47 days after each quarter end. Therefore, for the quarter from 2017-09-30 to 2017-12-31, this quarter is the institutional filing period and during this quarter we collected funds' filings. At 2017-12-31, this is the actual time we built our portfolio. Overall, when we built copycat strategy, we need to gap one quarter to avoid look-over bias.

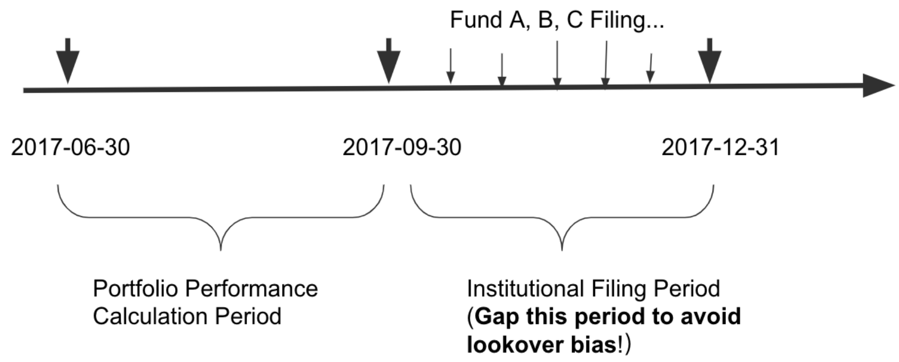


Figure 1: Copycat Strategy

3.6 Inflow

After calculating holding returns and total assets for each fund portfolio at each quarter end, we follow Sirri and Tufano (1998) [6] definition of mutual fund flows, and calculate net flows into institutional equity assets as the net growth as the net growth in assets. The definition of net inflow (or net outflow if negative) is the dollar value of buys minus sells from investors. As we have mentioned above, there are two types of buys: initiating buys and incremental trades. Similarly, two types of sells are terminating sells and partial sells. We summed up all buys and sells respectively and took the difference between these two numbers to get net inflows or outflows for each fund.

3.7 Turnover

As previous mentioned, we assumed that funds transactions happened after funds' filing dates, each funds current quarter stock holdings compared to past quarter has the stocks that are holing, buying and selling. We examined the performance of stocks held and traded by funds with varying levels of portfolio turnover.

$$Turnover_{k,t} = \frac{\min(\text{Buy}_{k,t}, \text{Sell}_{k,t})}{\text{TotalAssets}_{k,t}}, \tag{5}$$

Where $\text{Buy}_{k,t}/\text{Sell}_{k,t}$ is the total value of stock purchases / sales during quarter t by fund k, and $\text{TotalAssets}_{k,t}$ is the average total assets (market value of all holding positions) of fund k during quarter t. [7]

The following table gives several examples of quarterly fund performance information calculated using the methods mentioned above.

[6] Sirri and Tufano (1998)

[7] Fromhttps://www.jstor.org/stable/pdf/2676208.pdf?refreqid=excelsior%3A49ec0db088d3eb219d125dde4526117f

Period Start	Funds	CIK	Return	Total Initiating Buys (in dollars)	Total Terminating Sells (in dollars)	Total Incremental Trades (in dollars)	Total money Inflow (in dollars)	Total Assets of This quarter (in dollars)	Last Total Assets in Last Quarter (in dollars)	Turnover Ratio
6/30/2012	Indiana Trust & Investment Management CO	0001356407	-5.66%	$ 2,315,558.00	$ 5,218,816.00	$ 690,045.61	$ (2,213,212.39)	$ 103,727,306.00	$ 103,937,850.00	2.23%
6/30/2013	Hanson McClain, Inc.	0001555170	1.64%	$ 2,370,000.00	$ 536,994.00	$ (4,232,597.42)	$ (2,399,591.42)	$ 176,370,000.00	$ 177,908,294.00	0.30%
6/30/2014	SALEM CAPITAL MANAGEMENT INC	0001049648	4.26%	$ 200,000.00	$ 5,597,000.00	$ (6,371,536.26)	$ (11,768,536.26)	$ 222,014,000.00	$ 224,597,000.00	0.09%
6/30/2015	Quadrant Capital Management, LLC	0001615359	-2.59%	$ 5,080,000.00	$ 3,302,000.00	$ 801,444.66	$ 2,579,444.66	$ 116,625,000.00	$ 116,221,000.00	2.84%

Figure 2: Table of Performance Metrics Examples

In addition to the above aggregate statistics at fund level, our fund performance information also captured every trade an institution executed in one quarter. The following shows the volume of each incremental trade executed by the fund **Quadrant Capital Management**, **LLC** in the quarter end of 2015-06-30. The other plot shows the CUSIPs of securities newly added, de-listed and remained in the portfolio respectively.

```
res['2015-06-30']['0001615359']['Total hold positions']
```

	iCUSIP	iQTY_x	PRC_OLD	iQTY_y	position_changed
0	693475105	25025.0	91.4100	25825.0	-800.0
1	464287234	11750.0	42.2735	12250.0	-500.0
2	74005P104	8045.0	122.7900	8270.0	-225.0
3	17243V102	44519.0	43.0200	44719.0	-200.0
4	85254J102	26760.0	22.4000	25635.0	1125.0
5	464287630	48855.0	99.9259	47875.0	980.0
6	74340W103	20051.0	42.2300	19468.0	583.0

Figure 3: Example: Volume of Every Incremental Trade in This Quarter

```
hold ['004239109' '015271109' '026874784' '037833100' '075887109' '101121101'
 '166764100' '171340102' '17243V102' '22160K105' '229663109' '253868103'
 '268648102' '278642103' '293792107' '297178105' '30219G108' '30225T102'
 '32054K103' '369550108' '369604103' '375558103' '38141G104' '42217K106'
 '437076102' '437306103' '44107P104' '464287234' '464287465' '464287614'
 '464287630' '464287655' '543881106' '59522J103' '654106103' '655664100'
 '66987V109' '674599105' '68389X105' '693475105' '74005P104' '74144T108'
 '74340W103' '747525103' '74965L101' '756109104' '78464A698' '78464A870'
 '806857108' '81369Y407' '81369Y605' '828806109' '85254J102' '891894107'
 '89785L107' '913017109' '92826C839' '97717W851' '97717X701' 'G1151C101'
 'M22465104' 'N00985106']

buy ['46625H100', '268948106', '78467V848', '30231G102']

sell ['064149107', '25243Q205', '464288844']
```

Figure 4: Example: CUSIPs of Securities in the Portfolio (Grouped by different types of trade)

3.8 Limitations of Our Performance Measures

Note that our performance metrics were not able to capture institutional trades with frequencies higher than the holdings reporting frequency. For example, if some funds bought some stocks and then they closed all their positions on these stocks within the same quarter, it is impossible for us to capture this kind of trades solely from 13f data.

Furthermore, SEC does not require institutions to report their short positions and short-selling trades. Thus our fund portfolio performance metrics is merely based on the long portion of institutional holdings.

A controversial point in our calculation is our turnover measure. In Carhart (1997), the turnover ratio was defined as the minimum of buys and sells divided by average assets because the difference between purchases and sales is equal to the net flows of money from investors. However, Chen, Jegadeesh and Wermers (2000) did not think such definition captures money inflows or outflows of investors.

Another limitation in our data is that we include only the original 13F filings, and do not reflect subsequent amendments with corrections or additions to the original 13F holding data. This could be a huge problem if later amendments of previously confidentially treated securities make a lot of changes to the original filing, which means the original filings we used in our data could be very inaccurate. Also, institutions could conceal their most valuable positions from initial filings and thus we might lose a huge amount of useful information using only the original filings.

3.9 Data Filter

In chapter 2, we cleaned the price and fund data. Here, We applied the following filters to choose our fund data set from cleaned data:

1. Turnover: quarterly turnover < median of quarterly turnover
2. Survival time: a fund must survive for at least 2 consecutive quarters
3. Volatility: quarterly volatility < median of quarterly volatility

The reason we applied the first turnover filter is because 13F data is collected on a quarterly basis. Therefore, if the holding of a fund has changed dramatically after it discloses its holding, the holdings information contains no information value. We calculated the turnover for every fund in every quarter, and required the turnover rate to be smaller than the median of quarterly turnover. For the seconds filter, we need at least 2 quarters' data to calculate returns. For the third filter, we are more risky-reverse and we want to keep our volatility low. Therefore, for each quarter, we only kept those funds whose volatility is smaller than the median of quarterly volatility.

After applying all of those filters, we reduced the number of funds to around 400.

4 Stock-level Predicting Model

4.1 Introduction

In this project, we used two strategies to build the portfolio: stock-level predicting model and fund-level Copycat strategy. Since 13f fund data only has long data without short data, for both strategies, we only considered long only strategy in our portfolio construction. Copycat strategies will be described in Section 5.

In this section, we will focus on stock-level prediction models. They are classification models that use features extracted from cleaned and filtered 13f and price data to predict future stock price moving directions. For future study, if more features with high predicting power were constructed, we could have built regression models that predict stock prices instead of stock movement then apply Markowitz Portfolio Theory to build the portfolio. However, for preliminary tests, portfolio built from regression models did not perform well. Therefore, we instead used classification models and assigned equal weights to all stocks that predicted to have positive future returns.

4.2 Feature Description

As described in Section 3.1 and 3.2, we split each fund into 10 industry sub-funds or 9 size/-value sub-funds. Each sub-funds corresponding to cluster w only includes stocks classified as cluster w. For example, fund f held $s_1, \cdots, s_{20}$. By calculating correlation between $s_1, ...s_{20}$ and 10 industry ETFs, we classify s_1, s_2 into industry cluster 1, s_3, s_4 into industry cluster 2... Then we split fund f into sub-funds $f_1, ...f_{10}$ where f_1 only includes weights of stocks s_1, s_2, i.e. stocks classified into industry cluster 1, and so on. Similarly, by classifying stocks into 9 size/value clusters, we can split fund f into 9 sub-funds. From the perspective of each stock s, s must belong to one of the 10 industry clusters and one of the 9 size/value clusters; if fund f held stock s, then there must be two corresponding sub-funds split independently from fund f that held stock s.

After splitting funds into sub-funds, we define "good funds" within each cluster for each classification method in the following way. First, we calculated the correlation between each sub-fund and each ETF and sort the correlations in descending order. Large correlation means the fund specializes in this sector. We selected sub-funds with $t = 3$ highest correlation. Second, since we are not interested in funds that only buy industry ETF nor funds buy stocks in the industry that performed poorly, we calculated excess return of each sub-fund over each ETF and we only kept those sub-funds with positive excess return. From first and second steps, only sub-funds with high correlation with ETF and beat ETF could be selected for further steps of building features. These sub-funds are defined as "good funds".

Below are the specific descriptions of each feature we constructed.

- x_1: number of funds that held stock s in past quarter
- x_2: which industry cluster stock s belongs to. Definition and classification methodology of industry clusters were described in Section 3.1. For modeling purpose, we used LabelEncoder to transfer String categorical variables to numerical categorical values.
- x_3: which size/value cluster stock s belongs to. Definition and classification methodology of size/value clusters were described in Section 3.2. We used LabelEncoder to transfer strings into numerical values.
- x_4: within number of all funds held stock s, the percentage of number of good funds defined by industry cluster that held stock s
- x_5: mean weights of good funds defined by industry cluster that held stock s
- x_6: mean value of good funds past quarterly return defined by industry cluster
- x_7: standard deviation of good funds past quarterly return defined by industry cluster
- x_8: excess return of good funds past quarterly return defined by industry cluster over each industry ETF
- x_9: correlation between good funds past quarterly return defined by industry cluster and each industry ETF
- x_{10}: within number of all funds held stock s, the percentage of number of good funds defined by size/value cluster that held stock s
- x_{11}: mean weights of good funds defined by size/value cluster that held stock s
- x_{12}: mean value of good funds past quarterly return defined by size/value cluster
- x_{13}: standard deviation of good funds past quarterly return defined by size/value cluster
- x_{14}: excess return of good funds past quarterly return defined by industry cluster over each size/value ETFs
- x_{15}: correlation between good funds past quarterly return defined by industry cluster and each size/value ETFs
- $x_{16}/x_{17}/x_{18}$: 30/60/90 days historical returns
- x_{19}: idiosyncratic risk of funds that are currently holding stock s as described in Section 3.4
- x_{20}: fund holding this stock number change in market value from past 2 quarters to past quarter

- x_{21}: fund holding this stock number change in quantity from past 2 quarters to past quarter

The figure below is an example of what our feature looks like.

CUSIP	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
001055102	71	9	7	1.549296	0.012234	0.098040	2.952420	0.030061	0.982902	1.830986	0.012372	0.090498	2.848873
00206R102	281	0	2	0.014235	0.014412	0.127972	1.902101	0.067613	0.903537	0.053381	0.029846	0.089294	2.259552
002824100	242	1	4	0.016529	0.021457	0.095949	2.390177	0.038575	0.953880	0.330579	0.019241	0.088905	2.902128
00287Y109	241	9	7	1.236515	0.013822	0.094000	2.922470	0.027100	0.981833	1.775934	0.013615	0.090319	3.096439
00724F101	48	5	6	0.125000	0.011145	0.077532	3.119693	0.015444	0.961862	0.083333	0.002270	0.070687	2.928768
00817Y108	35	5	6	0.142857	0.004927	0.077835	2.064308	0.013545	0.971709	0.085714	0.010507	0.096272	1.624710
008252108	11	4	2	1.090909	0.002300	0.090896	3.338684	0.025827	0.987645	0.272727	0.010915	0.103636	4.668431
00912X302	10	4	1	2.400000	0.071901	0.090683	3.078123	0.039856	0.961926	2.000000	0.056876	0.094693	4.687565
009158106	75	5	6	0.080000	0.032086	0.098394	4.606394	0.024713	0.976768	0.026667	0.096163	0.076916	4.266490
00971T101	9	9	7	2.222222	0.017974	0.093497	2.807294	0.031759	0.976835	1.222222	0.017780	0.107238	3.774464

Figure 5: Partial Example with Feature Dataset

After building the features, we corrected outliers by defining all values greater than 3 standard deviations away from the mean as outliers and assign $mean + 3 \times std$ or $mean - 3 \times std$ as their values. We looked at each feature and analyzed its importance by three measurements:

- A scatter plot of future quarterly return versus each feature and see if there is any correlation.
- A linear regression between future quarterly return versus each feature and read its coefficient and p-value. We set significance level of 1% and define feature to be insignificant if p-value is greater than significance level.
- A box plot between future quarterly return versus each feature by 10 deciles. The box plot shows the trend more clearly than scatter plot.

The results of 2017/06/30 are shown below. The results from other dates are similar in terms of which features are significant. More details are attached in the Appendix. We did not plot features $x2, x3$ because they are sector categorical variables whose values do not have numerical meanings, but they should be included in the final model because other features are based on sectors. The subplot to the left is the scatter plot together with the linear line and coefficient with p-value. The subplot to the right is the box plot.

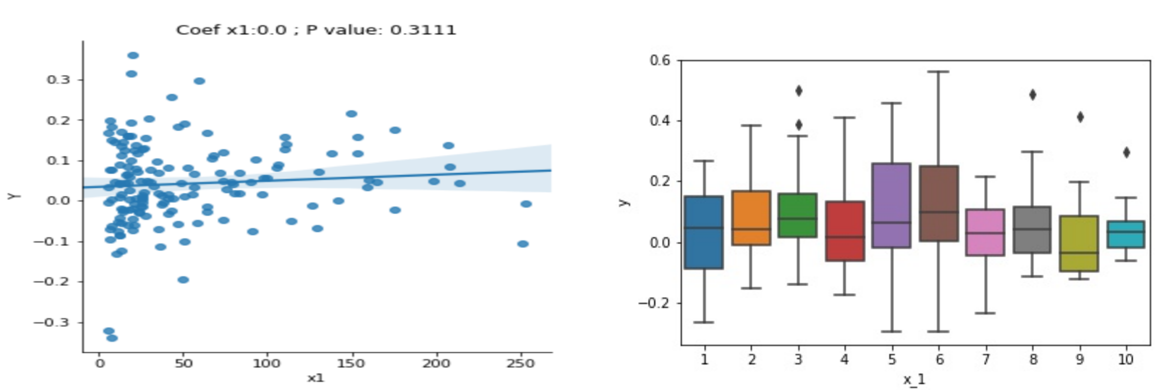


Figure 6: Feature 1 scatter plot (left) and box plot of deciles (right)

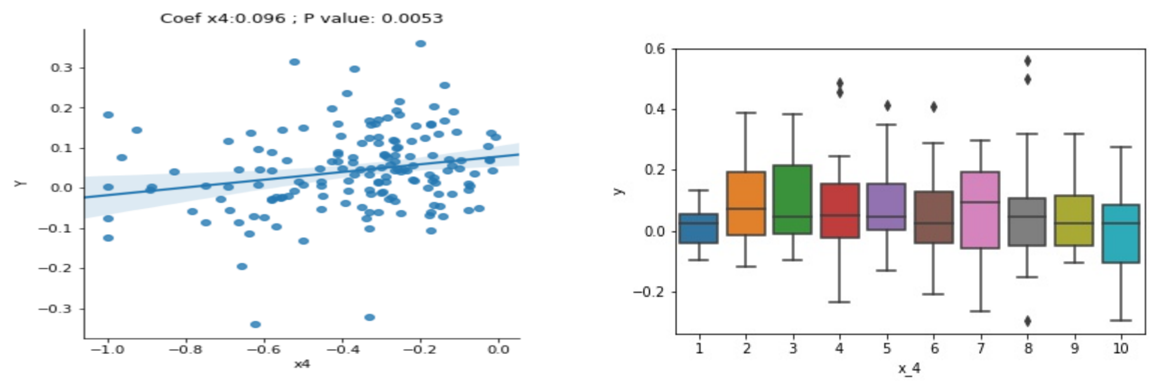


Figure 7: Feature 4 scatter plot (left) and box plot of deciles (right)

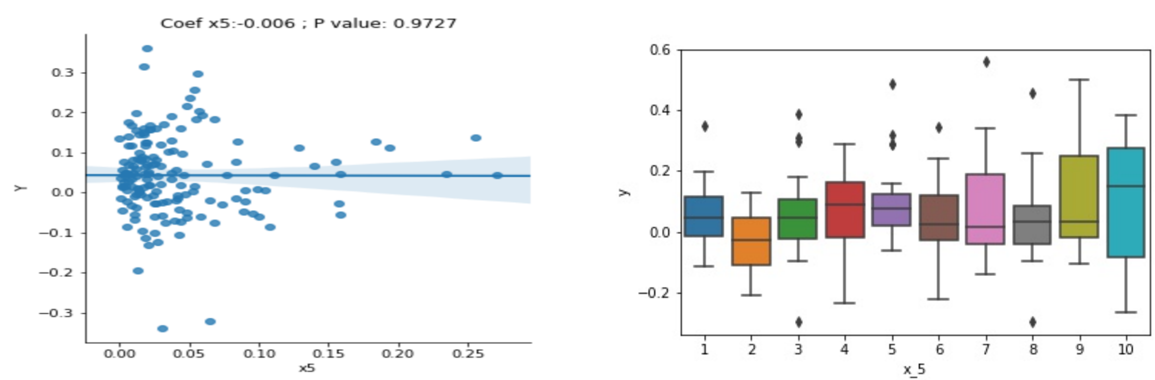


Figure 8: Feature 5 scatter plot (left) and box plot of deciles (right)

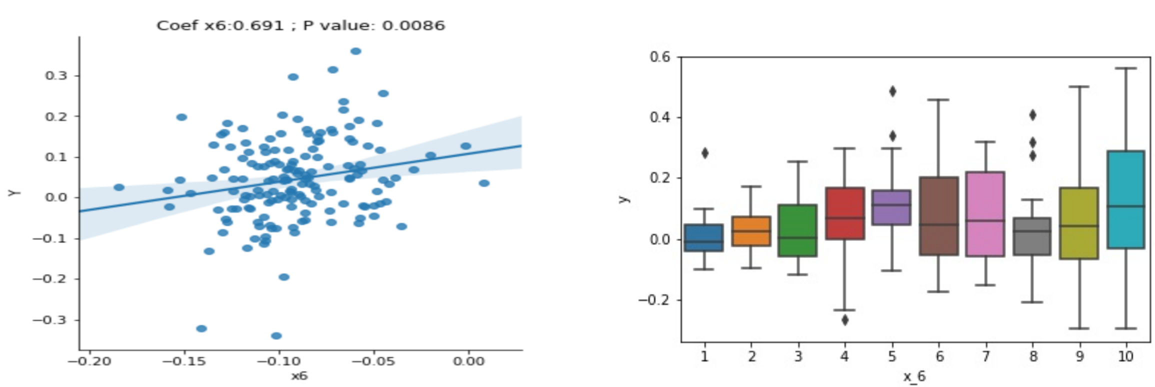


Figure 9: Feature 6 scatter plot (left) and box plot of deciles (right)

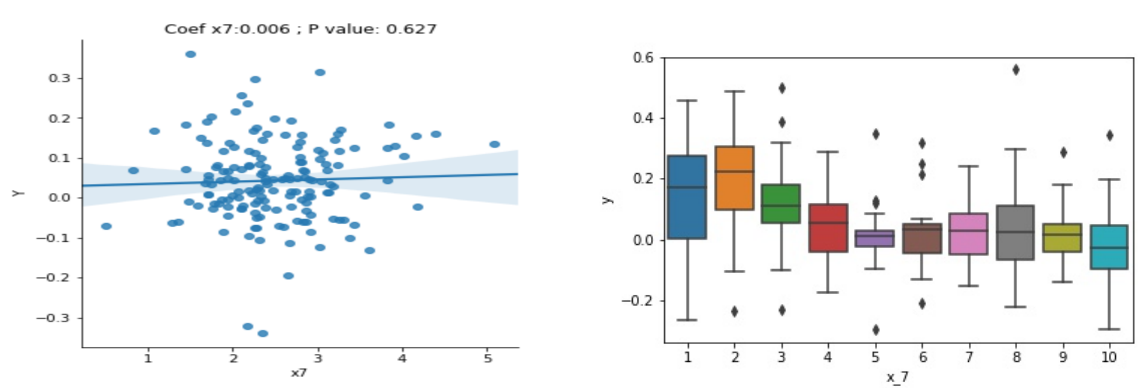


Figure 10: Feature 7 scatter plot (left) and box plot of deciles (right)

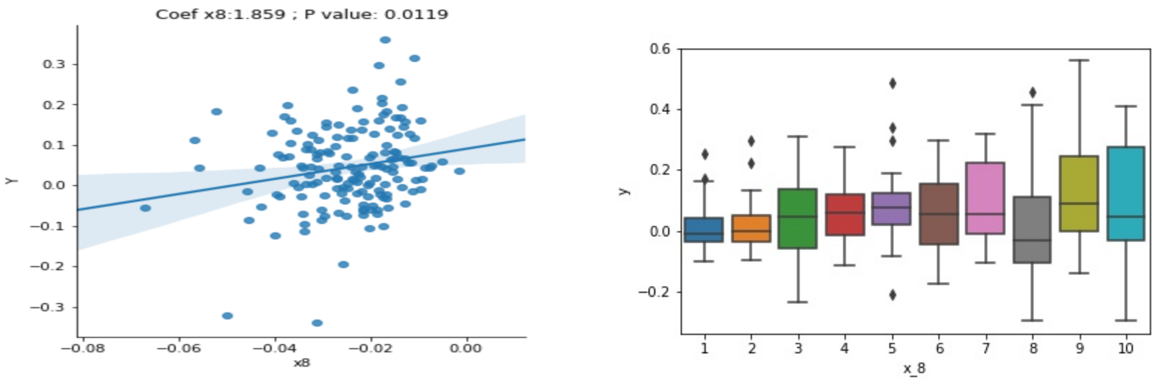


Figure 11: Feature 8 scatter plot (left) and box plot of deciles (right)

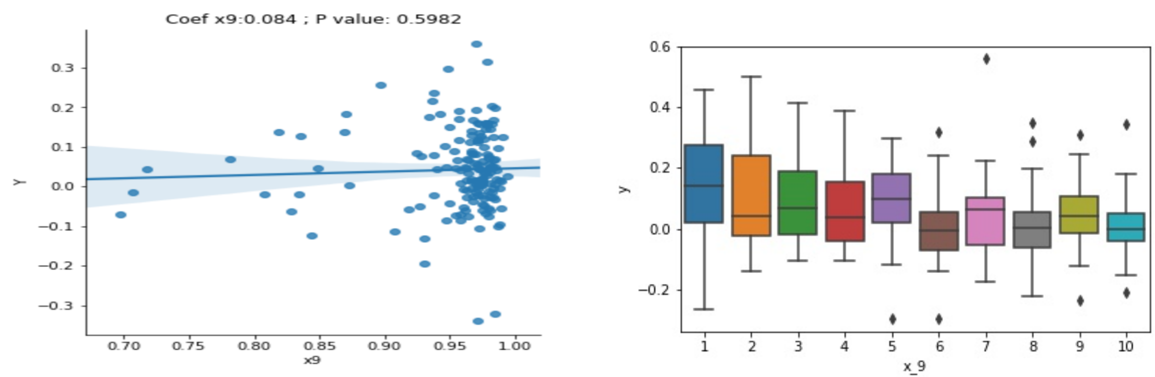


Figure 12: Feature 9 scatter plot (left) and box plot of deciles (right)

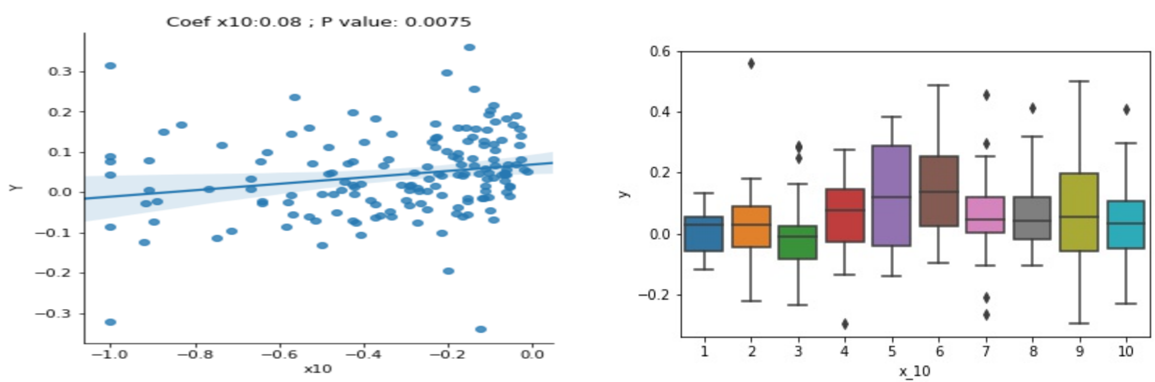


Figure 13: Feature 10 scatter plot (left) and box plot of deciles (right)

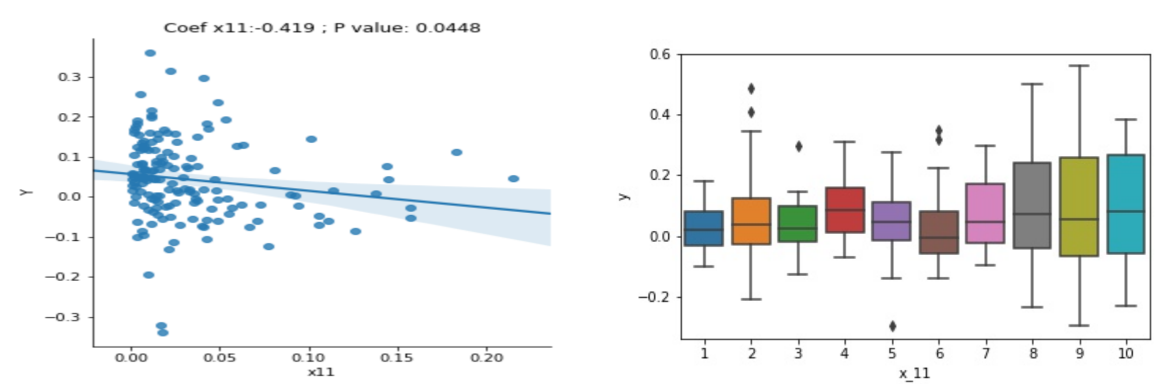


Figure 14: Feature 11 scatter plot (left) and box plot of deciles (right)

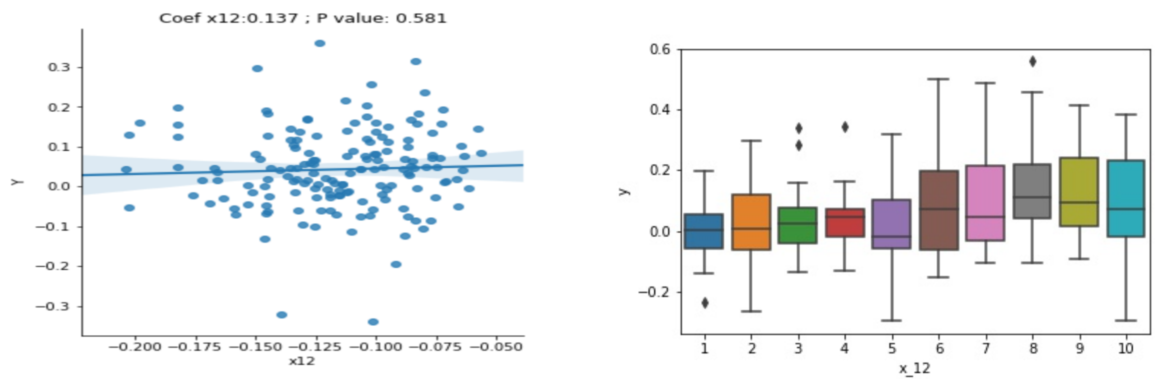


Figure 15: Feature 12 scatter plot (left) and box plot of deciles (right)

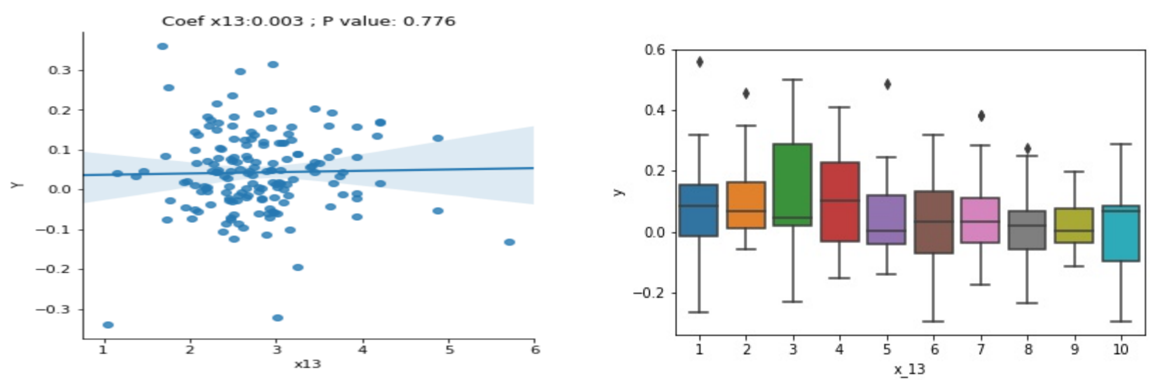


Figure 16: Feature 13 scatter plot (left) and box plot of deciles (right)

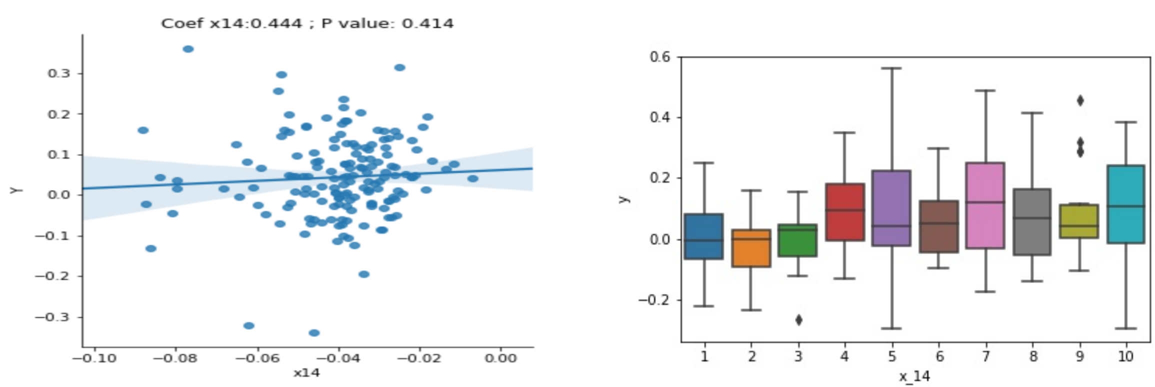


Figure 17: Feature 14 scatter plot (left) and box plot of deciles (right)

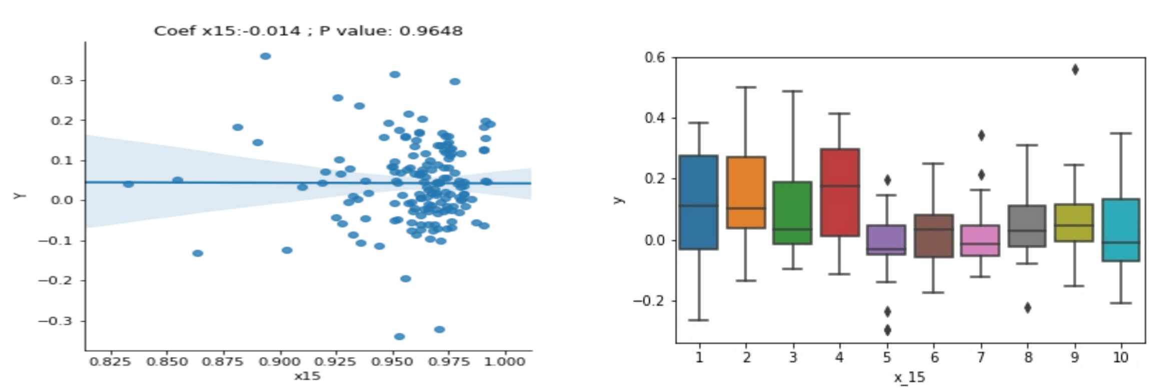


Figure 18: Feature 15 scatter plot (left) and box plot of deciles (right)

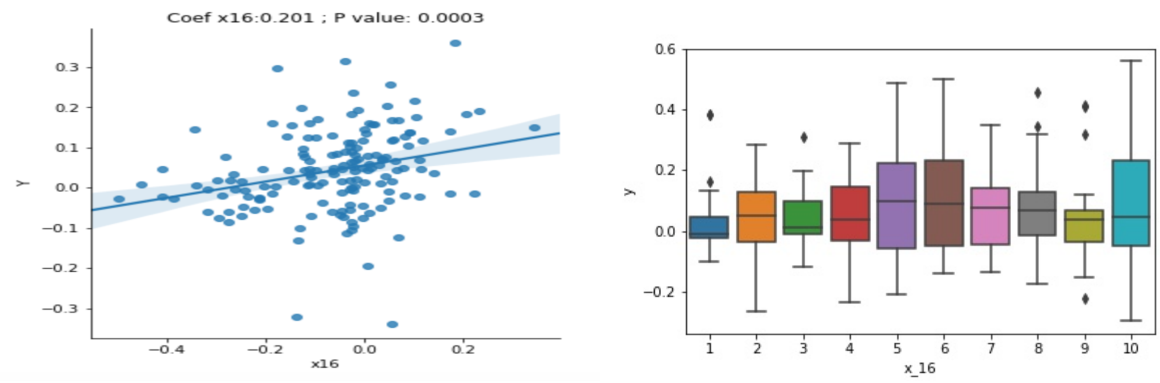


Figure 19: Feature 16 scatter plot (left) and box plot of deciles (right)

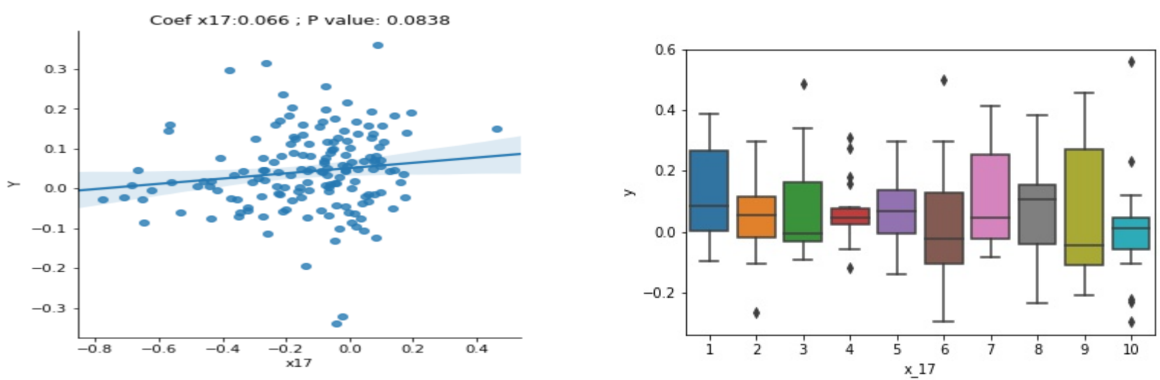


Figure 20: Feature 17 scatter plot (left) and box plot of deciles (right)

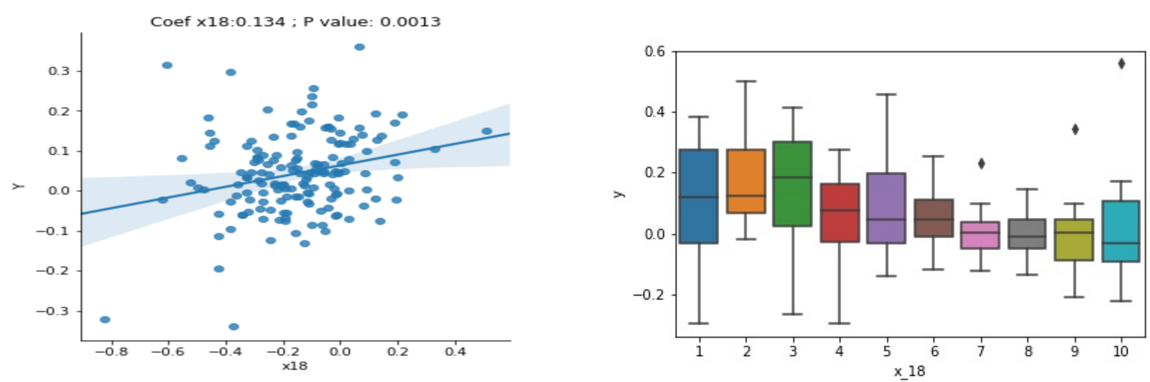


Figure 21: Feature 18 scatter plot (left) and box plot of deciles (right)

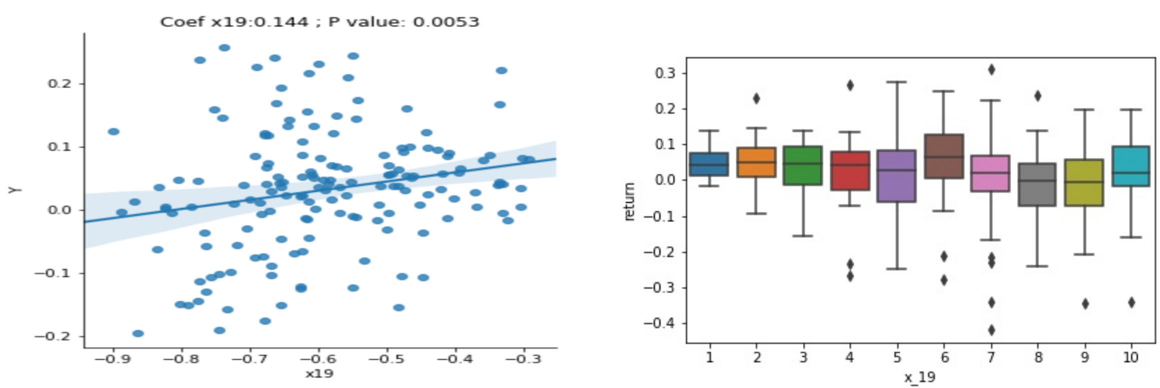


Figure 22: Feature 19 scatter plot (left) and box plot of deciles (right)

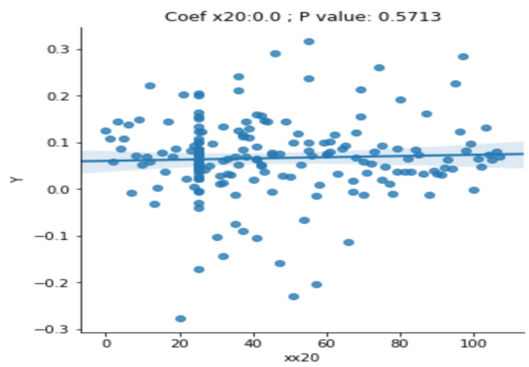


Figure 23: Feature 20 scatter plot (left) and box plot of deciles (right)

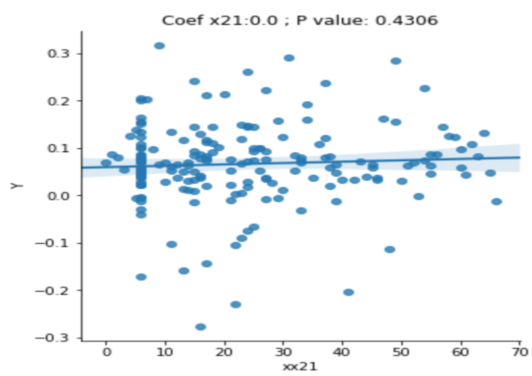


Figure 24: Feature 21 scatter plot

Note that, even though Feature 20 and 21 are categorical variable, they represents the number of funds that change the holding of a particular stock. This number can vary from 0 to the total number of funds of that quarter that we considered. Hence it makes sense to regard this categorical variable as a numeric variable. We have the following observations from the feature plots:

- From the scatter plot pvalue, we can see Feature 4,6,8,10,16,17,18,19 all give us pvalue smaller than 0.01, which indicates they are more significant in our predicting model.
- Feature 4, 10 scatter plots make sense because they are representing the good fund features. If more good funds hold a particular stock, it is more likely that this stock return to be more positive
- Feature 6, 8 are also expected. Since the better performance the fund is, the more likely that their funds held is performing good.
- Feature 16, 17, 18 scatter plots give consistent plots, and the higher the past return, the higher future return, which will also validate in Section 5.2.
- Feature 18 gives us a very consistent result in terms of scatter and box plot. The pvalue is relatively small, 0.0013, and there is a clear down-trend of return against feature percentage. To be more specific, from box plot when Feature18 increase its value gradually, the return will also decrease gradually, which is supported by the scatter plot.

- Feature 19 also gives a reasonable result. We know in 2017, the market is going uptrend, the higher idiosyncratic risk is, the more return that we can gain.
- Lastly, Feature 20 and 21 also make sense, since when more funds increase their holdings, regards to quantity or market value into some stocks, it is more likely that such stocks can give us good returns.

4.3 Modeling Methodology

After building features, we applied two stock pools, two different time series data structures, two Machine Learning models and two sets of features. We compared their return from 2013/06/30 to 2017/09/30.

First, we used all data we cleaned using methods described in Section 2.2. The table below shows number of unique funds and unique stocks in each quarter.

Table 4: Unique fund and stock number with all data pool

Time	# Unique funds	# Unique stocks
2013, 6, 30	949	4279
2013, 9, 30	307	4147
2013, 12, 31	360	4392
2014, 3, 31	381	4700
2014, 6, 30	465	5297
2014, 9, 30	476	5538
2014, 12, 31	279	4160
2015, 3, 31	375	4872
2015, 6, 30	442	5401
2015, 9, 30	355	4743
2015, 12, 31	380	5011
2016, 3, 31	409	5142
2016, 6, 30	468	5250
2016, 9, 30	385	4760
2016, 12, 31	354	4446
2017, 3, 31	535	5477
2017, 6, 30	457	4956
2017, 9, 30	550	5619

As will be explained later in Section 4.5, in our final portfolio, after we predict the stock returns, we selected 200 stocks to include in our portfolio ranking by trading volume. As a comparison, we built another set of features using stock pool of S&P500 stocks. We defined

SP500 stocks as all stocks that have made into SP500 lists from 2013 to 2018, which includes 847 stocks in total. The table below shows number of unique funds and unique stocks in each quarter.

Table 5: Unique fund and stock number with S&P 500 stock pool

Time	# Unique funds	# Unique stocks
2013, 6, 30	938	602
2013, 9, 30	306	600
2013, 12, 31	360	603
2014, 3, 31	375	603
2014, 6, 30	463	605
2014, 9, 30	470	603
2014, 12, 31	269	611
2015, 3, 31	370	612
2015, 6, 30	447	619
2015, 9, 30	453	619
2015, 12, 31	277	621
2016, 3, 31	398	623
2016, 6, 30	466	625
2016, 9, 30	378	619
2016, 12, 31	353	624
2017, 3, 31	516	619
2017, 6, 30	455	622
2017, 9, 30	538	625

Second, we compared two time series data structures. They differ from each other by the amount of data used to train the model: the first one used past quarterly data to train the model; the second one used all available historical data to train the model. In other words, the first structure used a moving window and the second structure used an expanded window. Both approaches predicted future quarterly return using feature from current quarter and model trained from past quarter or past all quarters data. The moving window approach has the advantage that it only looks back one period so that there is not much noise in the feature space from historical data. The expanded window approach has the advantage that there is much more data in the training set so that over-fitting problem could be possibly mitigated.

Third, we compared two Machine Learning models: Logistic Regression and Extreme Gradient Boosting Method (XGBoost). Logistic Regression is a widely used classification model with the following formulation:

$$log\frac{P(y=1)}{1-P(y=1)} = \alpha + \sum_{i=0}^{n} \beta_i * x_i$$

The biggest advantage of Logistic Regression is that due to its output format, $log\frac{P(y=1)}{1-P(y=1)} \in [0,1]$, its predicting results can be interpreted as probabilities. However, the Logistic Regression assumes linear relationship between dependent and independent variables so that it cannot capture nonlinear relationships. It additionally assumes that independent variables are not correlated with each other so that it does not capture any interactions among independent variables. Nevertheless, features we used in our model like $x11$ and $x12$ can be correlated with each other and taking interaction into account would increase the model prediction power.

Therefore, as a second attempt, we used XGBoost models. XGBoost is a recently widely used Machine Learning algorithm that is well-known for its high predicting power. It is one of Gradient Boosting methods, which are based on Decision Trees and use boosting method to ensemble multiple trees with weak predicting powers to generate a final tree with strong predicting power. Below is a general framework of a gradient boosting algorithm:

Input: training set $(x_i, y_i)_{i=1}^n$, a differential loss function L(y,F(x)), number of iterations M.

Algorithm:

1. Initialize model with a constant value: $F_0(x) = argmin_\lambda \sum_{i=1}^n L(y_i, \lambda)$
2. For m = 1 to M:
   (a) Compute so-called pseudo-residuals: $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]$ for i = 1, ..., n
   (b) Fit a base learner $h_m(x)$ to pseudo-residuals, i.e. train it using the training set $(x_i, r_{i,m})_{i=1}^n$
   (c) Compute the multiplier λ_m by solving the following one-dimensional optimization problem:
   $$\lambda_m = argmin_\lambda \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \lambda h_m(x_i) \tag{6}$$
   (d) Update the model: $F_m(x) = F_{m-1}(x) + \lambda_m h_m(x)$
3. Output $F_M(x)$

Fourthly, we compared models with all features versus models with only significant features as described in the previous subsection. In other words, we compared model with features $x1...x20$ and model with features $x4, x6, x8, x10, x11, x16, x17, x18, x19, x20$.

4.4 Portfolio Construction

After we used various methodologies and built different stock-level prediction model, we constructed portfolios based on the prediction model.

Since we want to build a long-only portfolio, with stock-level predictions, we first subset all stocks with positive predicted returns. Then, we limited our portfolio to include at most 200 stocks because we did not want our portfolio to be too diverse. We chose 200 stocks according to their trading volume since the more the average trading volume in the past quarter, the more likely the stock will outperform in the next quarter. Finally, we assigned equal weights to all stocks selected.

Further analysis of portfolios comparison are in Section 6.2.

5 Fund Copycat Strategy

5.1 Motivation

Recall that Form 13f data from SEC is a quarterly report filled by institutional asset managers about their of equity holdings under management. Since the SEC 13F fillings data can be utilized to track the positions of investors' portfolio, it is natural to think about an investment strategy that captures the intelligent ideas of fund managers, which is introduced as "Copycat" strategy in this section. This work tries to create a Copycat strategy that replicates the equities holdings based on disclosed information from 13f data over the quarter right after filing period quarter. We considered the quantities introduced in the section 3 such as return, inflow and turnover, etc. as indicator metrics to rank, select and replicate funds accordingly.

5.2 Methodology Analysis

To construct Copycat strategy, we first explored the time series of funds' turnover and cash inflows, and computed their correlation with returns. As mentioned in Chapter 3.9, when trying to create funds Copycat strategy, the funds that have relative high turnover ratio are not preferred, because it would not be reasonable to copy those funds whose large proportion of holdings are frequently changed. Hence, the long only holding positions reported by institutional managers with high turnover ratio can be useless for Copycat strategy. After plotting the histogram and calculating the median of the turnovers, we knew that the median of the funds turnover is 0.016. To avoid the negative effects of high turnover on Copycat strategy, we further filtered the funds with turnover higher than 0.016, and only kept funds with turnover lower than median.

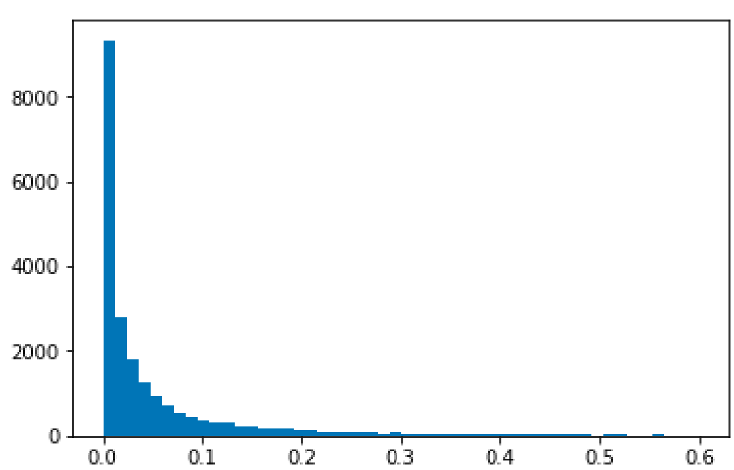


Figure 25: Histogram of Turnover

The histograms of the cash inflows and returns before the filter was applied are displayed below.

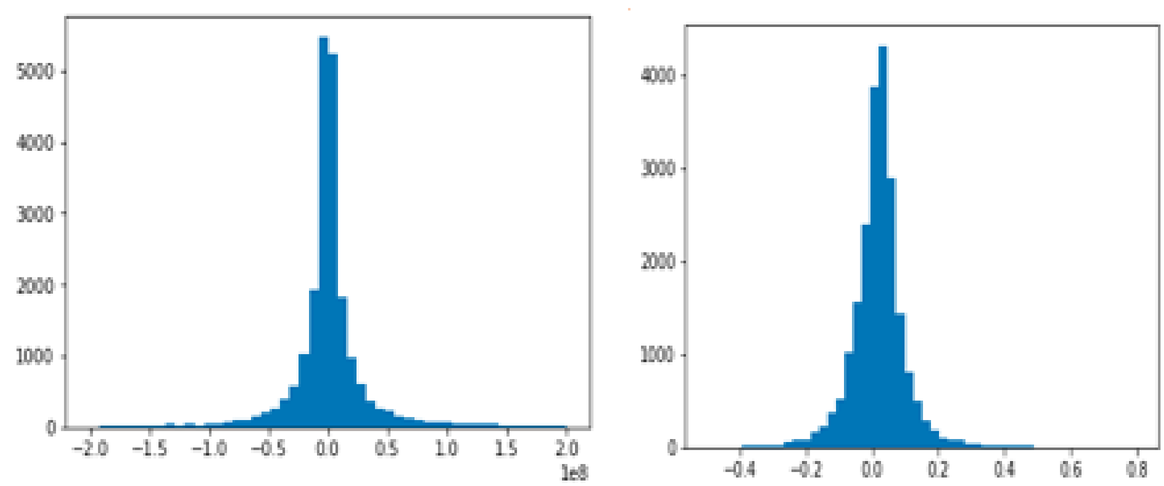


Figure 26: Cash Inflow (left) and Return (right) Distribution before Turnover Filter

Accordingly, the correlation of the cash inflows with the following period's return prior to the filter is 0.00378. Then, we re-plotted the distribution of cash inflows and returns. The histograms are as follows:

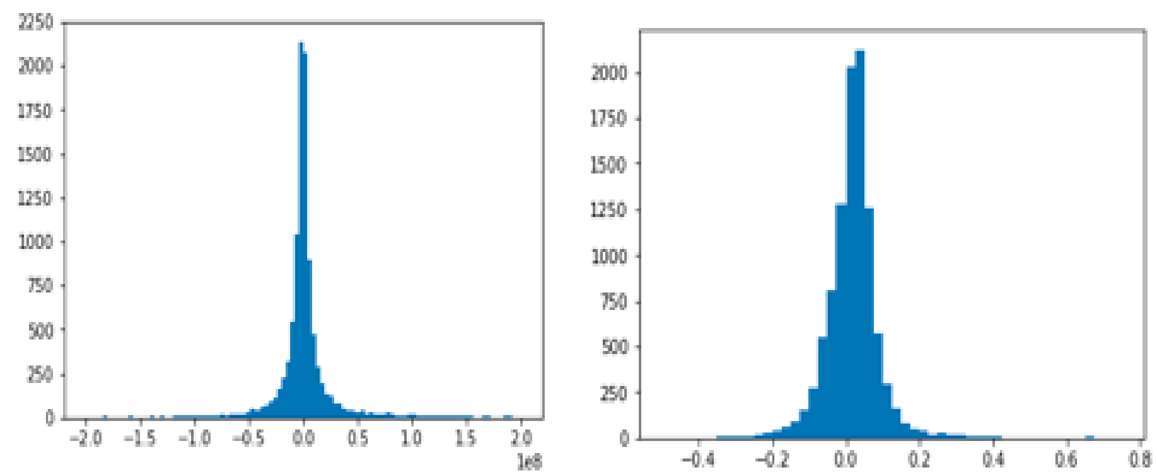


Figure 27: Cash Inflow (left) and Return (right) Distribution after Turnover Filter

With the turnover threshold 0.016 as a funds filter, the correlation of the cash inflows with the following period's return after filter is 0.007195 which significantly increased. Moreover, we also plotted the scatter plots of funds' returns in current quarter and funds' returns in next quarter. The visualization results are consistent with the numerical values of correlations: in general, funds may perform better in the next quarter period if it performed relatively well during the past quarter period. Thus, the correlation analysis of cash inflow and returns build stones and strong justification for our Copycat strategy.

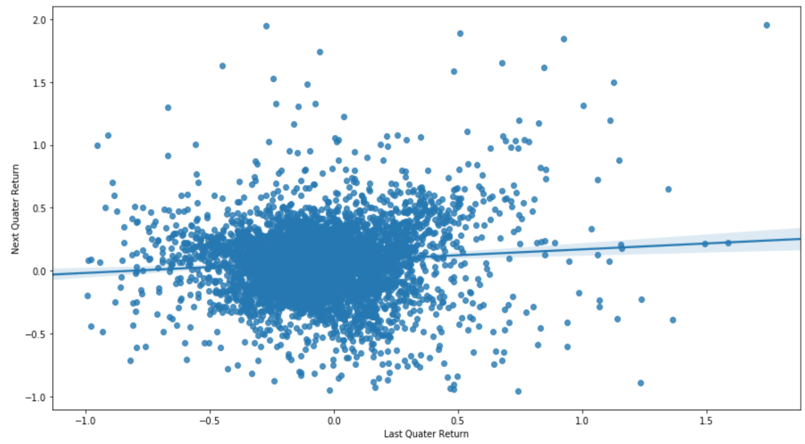
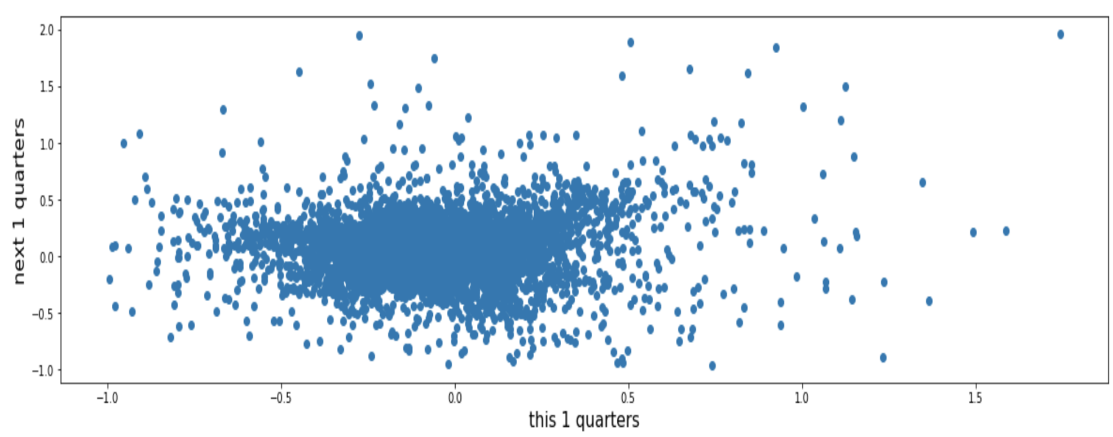
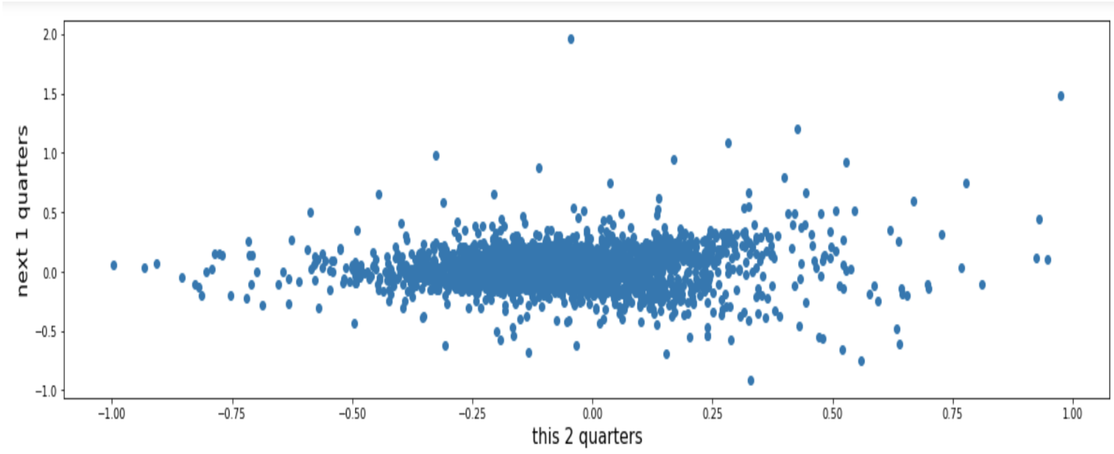


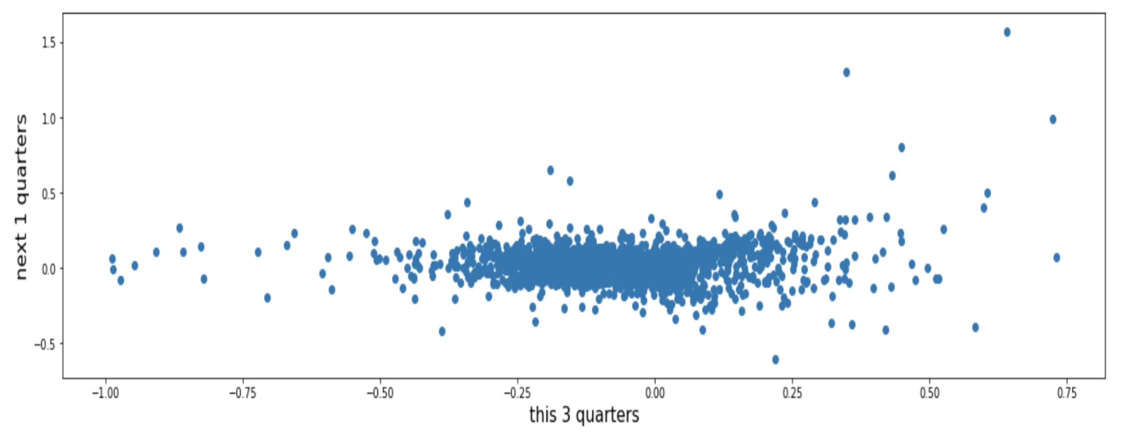
Figure 28: Next 1 Quarter Return vs This 1 Quarter Return

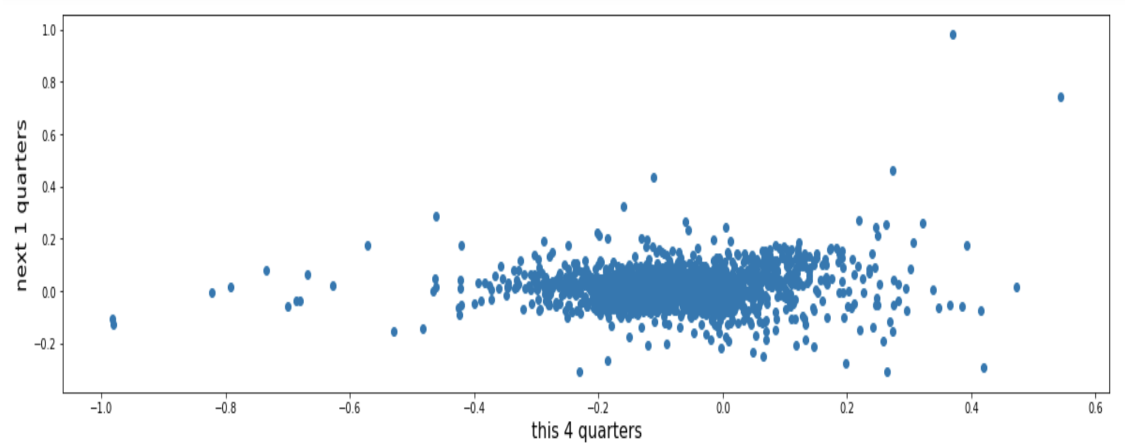
	Correlations	P-Value
Lookback 1 quarter return Vs. forward 1 quarter return	0.888	1.0
Lookback 2 quarters return Vs. forward 1 quarter return	0.0845	1.0
Lookback 3 quarters return Vs. forward 1 quarter return	0.0875	1.0
Lookback 4 quarters return Vs. forward 1 quarter return	0.1071	1.0
Lookback 5 quarters return Vs. forward 1 quarter return	0.0749	0.0057
Lookback 6 quarters return Vs. forward 1 quarter return	0.1083	0.0006
Lookback 7 quarters return Vs.forward 1 quarter return	0.2006	2.2048e-18
Lookback 8 quarters return Vs.forward 1 quarter return	0.2599	2.1334e-10

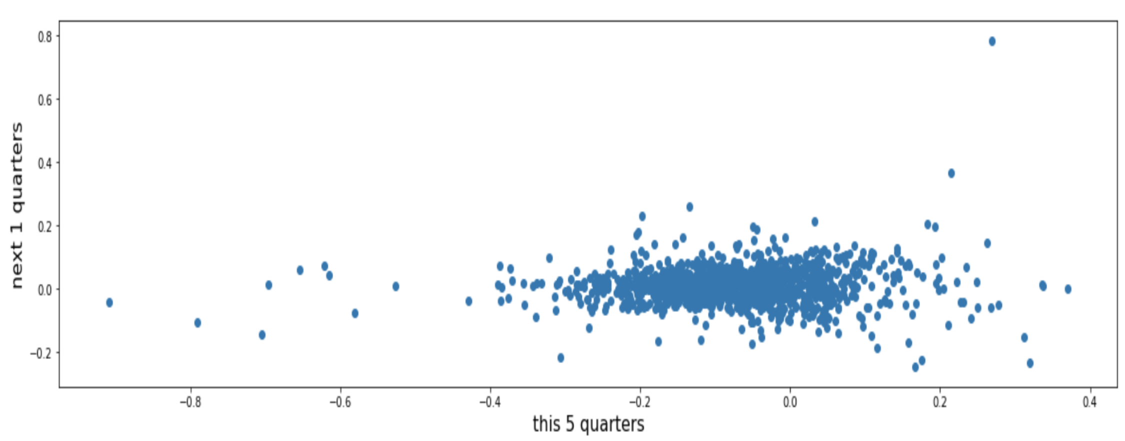
Figure 29: Correlation Table

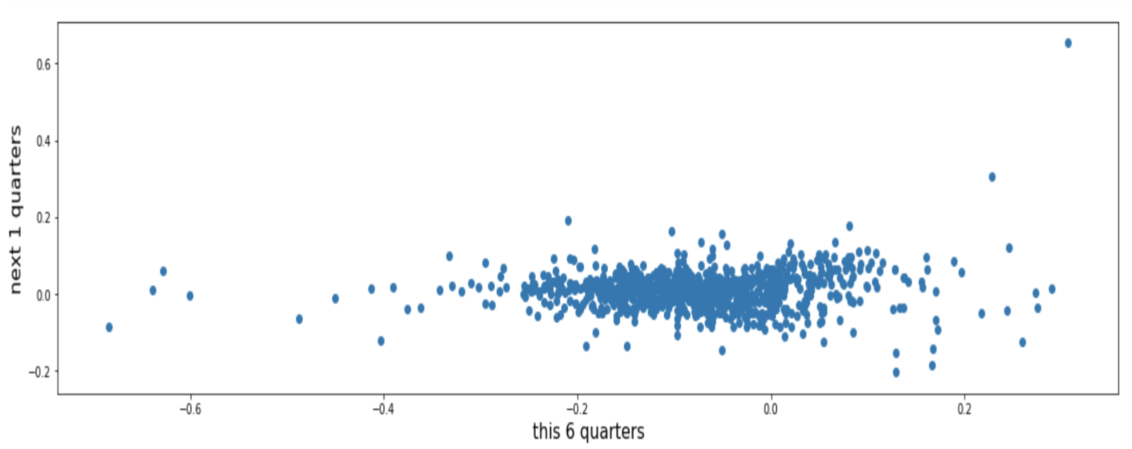


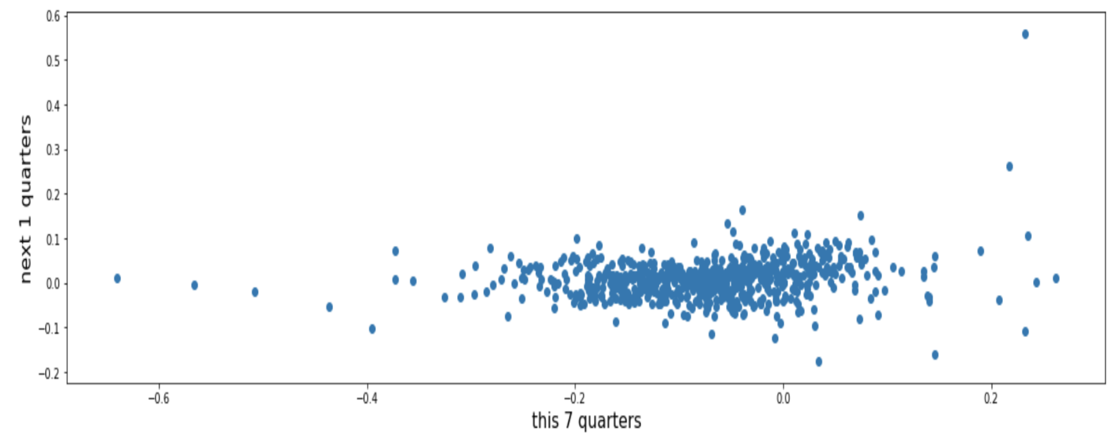



next 1 quarters
this 3 quarters


next 1 quarters
this 4 quarters


next 1 quarters
this 5 quarters


next 1 quarters
this 6 quarters



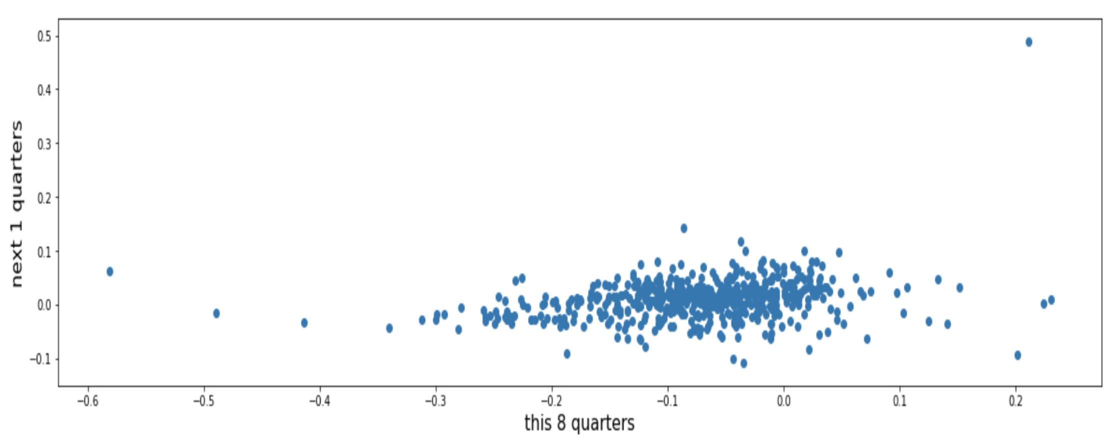
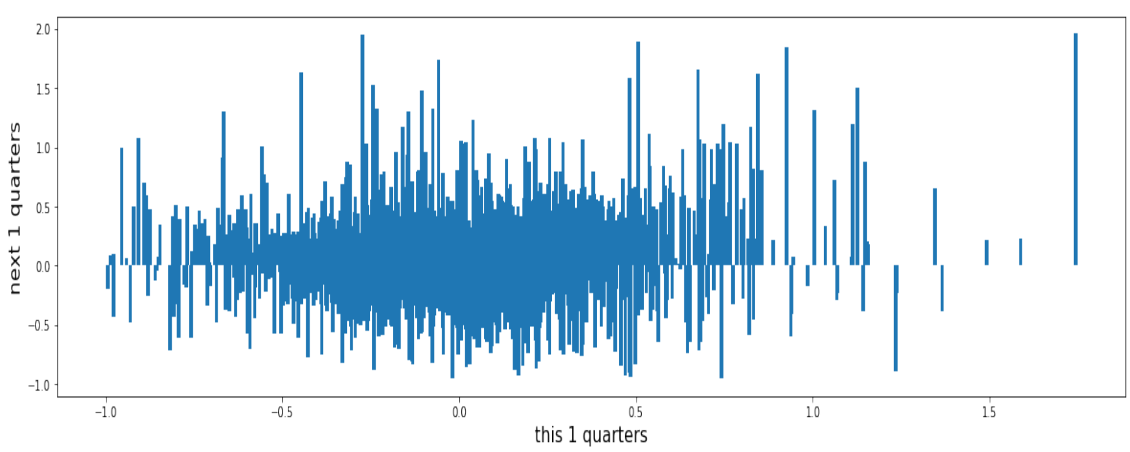
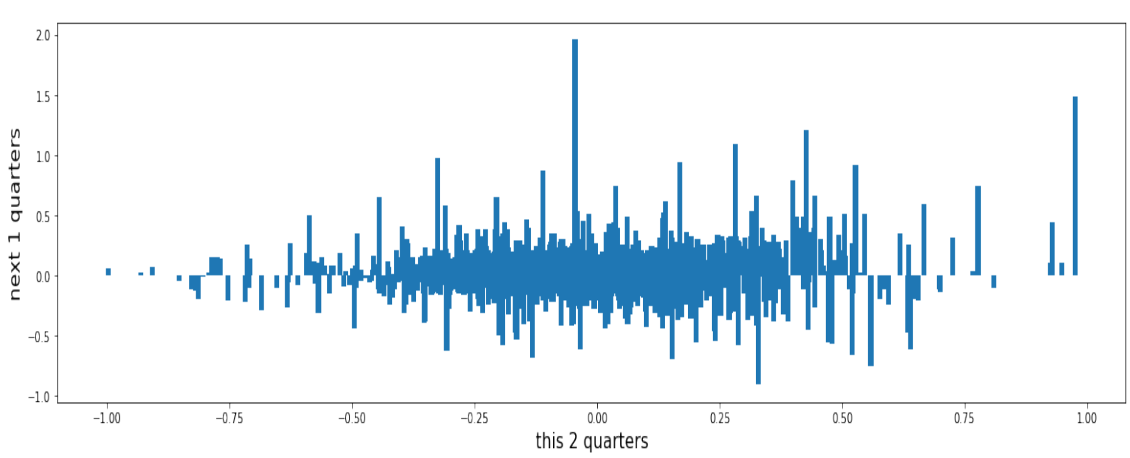
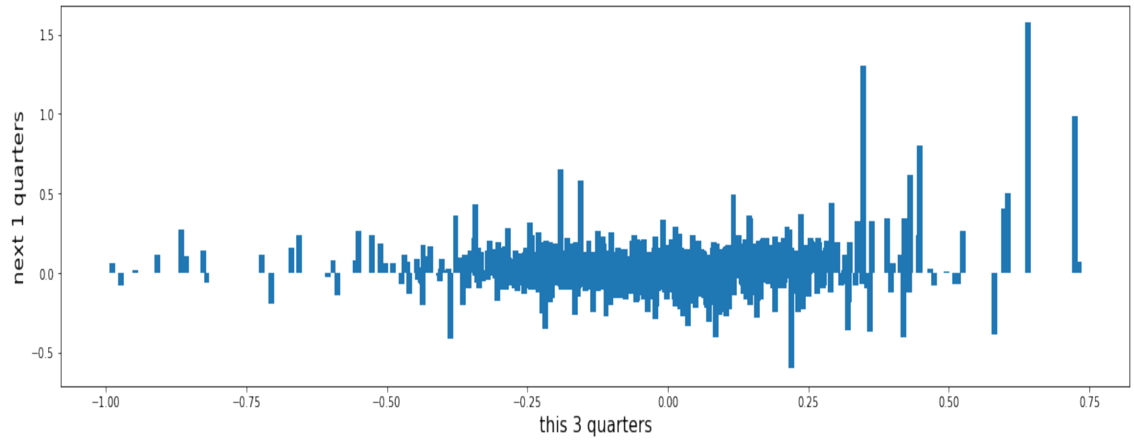
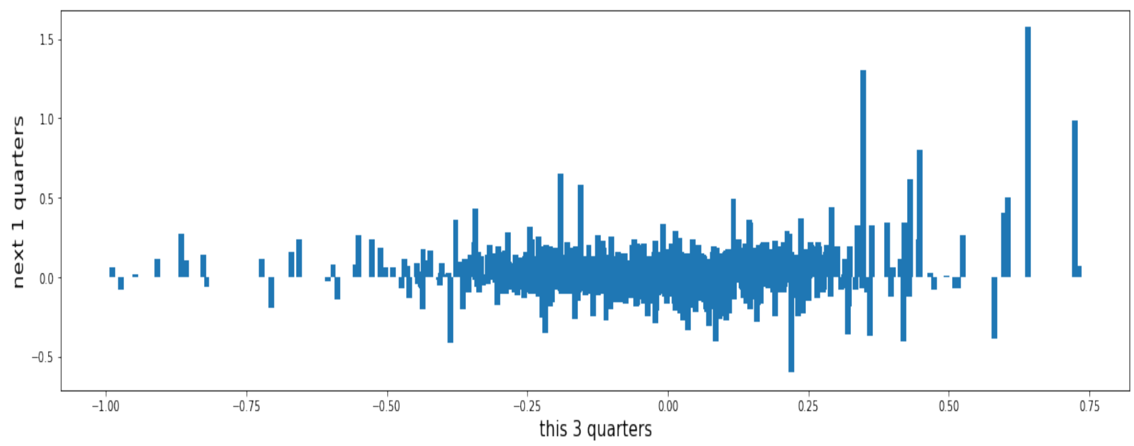


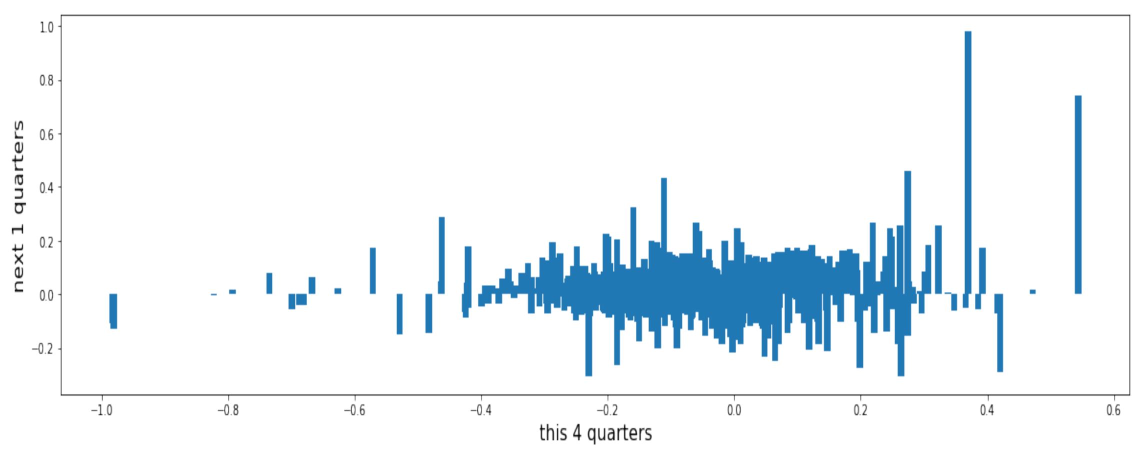
Figure 30: Correlations Between Historical Returns (with different lookback window size and Forward Returns)

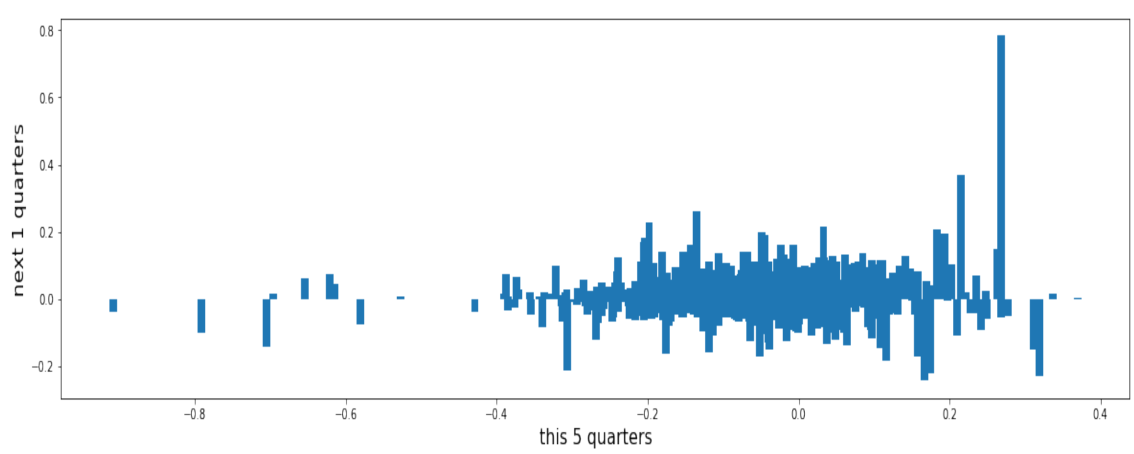





next 1 quarters
this 3 quarters
1.5
1.0
0.5
0.0
−0.5
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75


next 1 quarters
this 3 quarters
1.5
1.0
0.5
0.0
−0.5
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75


next 1 quarters
this 4 quarters
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6


next 1 quarters
this 5 quarters
0.8
0.6
0.4
0.2
0.0
−0.2
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4

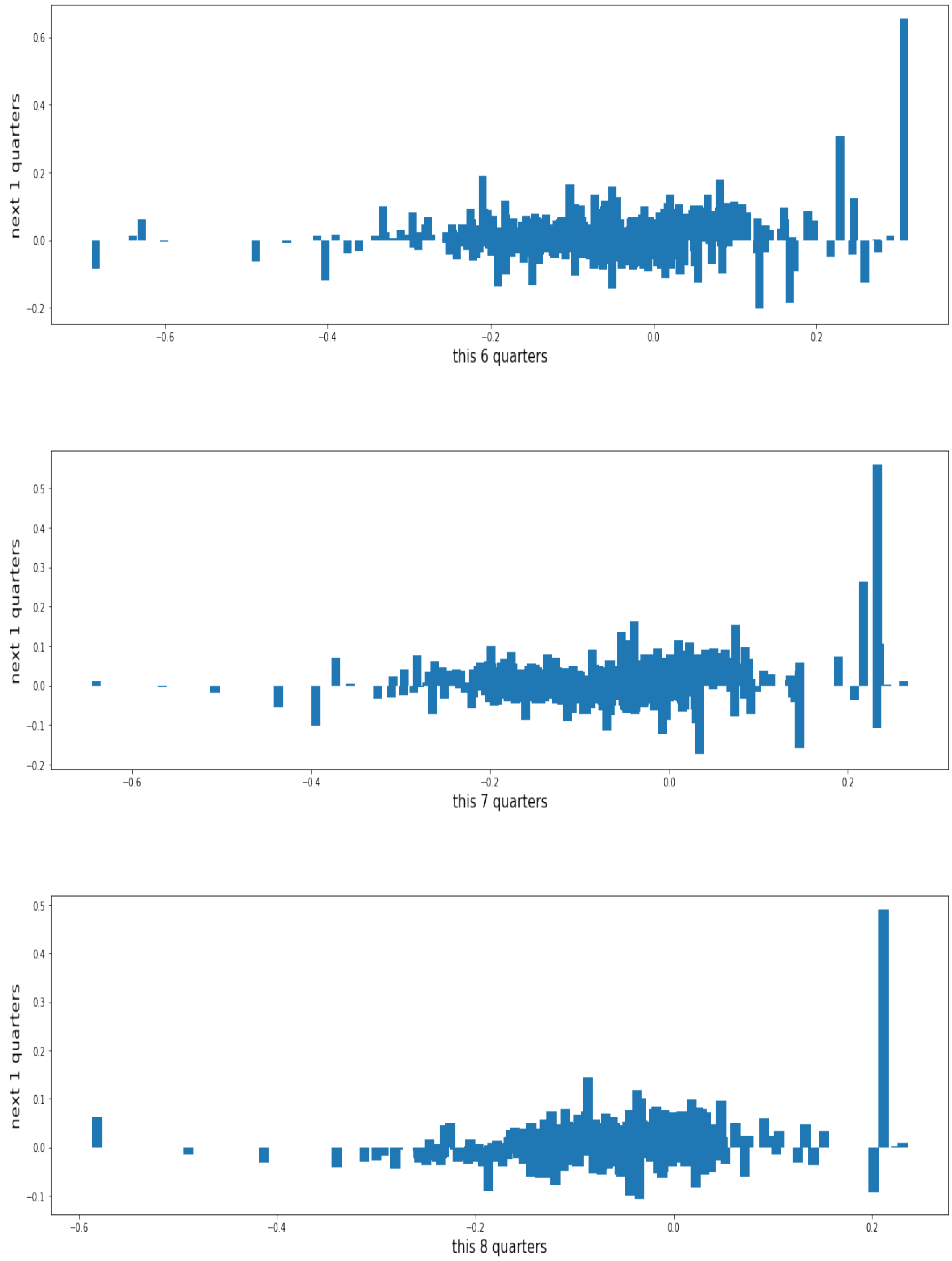


Figure 31: Correlations Between Historical Returns (with different lookback window size and Forward Returns)

The 8 plots of Figure 29 tell us that the longer the looking back window, the more funds with better return can get. To be more specific, for a particular fund, when we look back 8 quarters and replicate their holdings, the next quarter return in general, will be the highest. Especially, when looking back 8 quarters, the funds that perform very well at that time actually perform astonishingly well in the next quarter. This finding is also backed up by the Figure 27. We can see that there are indeed correlation around 0.26 between past 8 quarters return and next quarter return with the p-value extremely small. This finding is the building stone of our copycat strategy, where we focus on copycatting the funds which can give us the top returns in the past 8 quarters. We will quarterly copycat such funds and re-balance our holdings dynamically with the top funds.

6 Backtesting

6.1 Evaluation Metrics

We used 4 evaluation metrics to evaluate the performance of our portfolios:

1. Cumulative Return
2. Sharpe Ratio: $\frac{r_{annualized,P} - r_{annualized,rf}}{\sigma_{annualized,P}}$
3. Sortino Ratio: $\frac{r_{annualized,P} - r_{annualized,rf}}{\sigma_{annualized,downside}}$,
   where $\sigma_{annualized,downside} = k\sqrt{\frac{1}{N}\sum_{i=1}^{N} min(0,(r_i - r_{rf}))^2}$
4. Calmar Ratio: $\frac{r_{annualized,P} - r_{annualized,rf}}{\max_{DrawDowns}(r_{peak} - r_{trough})}$

The Sharpe ratio is a way to examine the performance of an investment by adjusting for its risk. The ratio measures the excess return (or risk premium) per unit of deviation in an investment asset or a trading strategy.

Besides the widely used cumulative returns and Sharpe Ratio. The Sortino ratio measures the risk-adjusted return of an investment asset, portfolio, or strategy. It is a modification of the Sharpe ratio but penalizes only those returns falling below a user-specified target or required rate of return, while the Sharpe ratio penalizes both upside and downside volatility equally. For the Calmar Ratio, unlike Sharpe Ratio and Sortino Ratio, it uses maximum draw downs as its risk proxy.

In Section 6.2 and 6.3, we show the backtest results for Stock Prediction Strategy and Copycat Strategy, which includes both graphical comparison of cumulative returns among models together with benchmark. We chose S&P 500 as out benchmark. In Section 6.4, we put together the best model from Stock Prediction Strategy and Copycat Strategy, compare their returns with benchmark, and additionally look at backtest statistics to measure the stability of each strategy.

6.2 Stock Prediction Strategy Backtest Results

We compared different stock prediction models in four dimension as explained in Section 4.4 by plotting the cumulative returns of portfolios built from each model from 2013/09/30 to 2017/12/31:

1. Data Pool: all stocks versus S&P 500 stocks
2. Data Structure: moving window versus expanding window

3. Feature Lists: all features versus significant features

4. Machine Learning Algorithm: logistic regression versus XGBoost

The following graph shows the comparison between all stocks versus S&P 500 stocks.

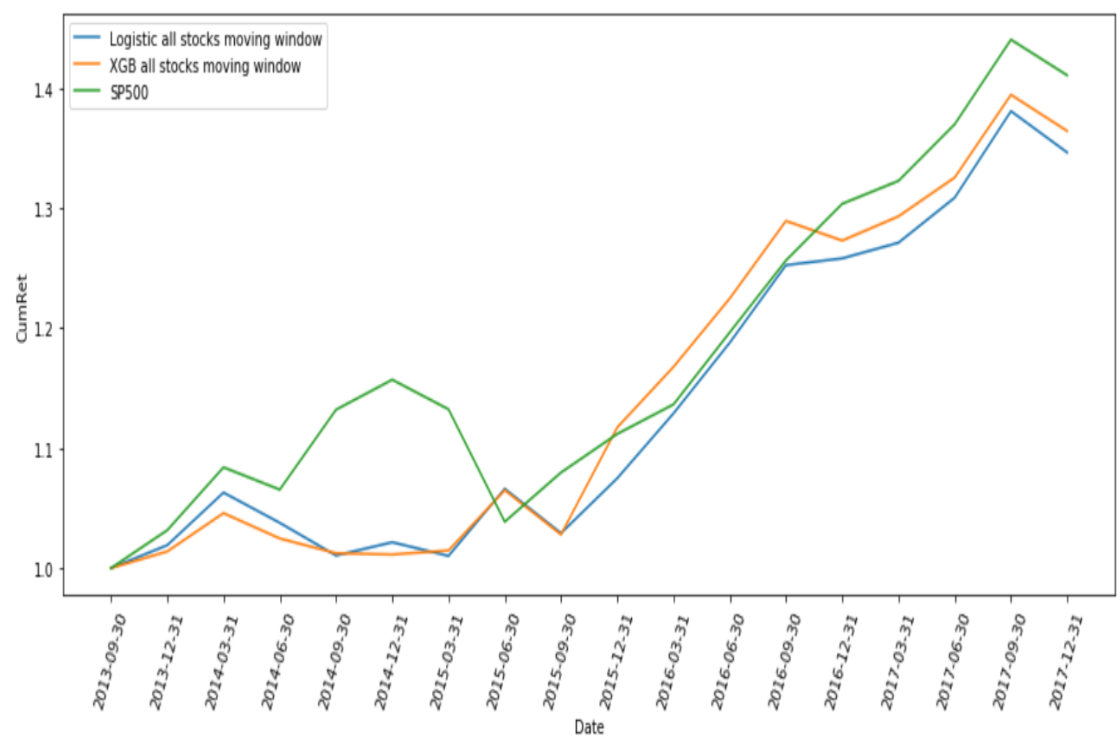


Figure 32: Stock Prediction Model All Stocks Moving Window Strategy Backtest Result

As we can see from the graph, XGBoost slightly outperformed Logistic Regression but neither strategy outperformed the benchmark. After we limited the stock pool to S&P 500, most models outperformed the benchmark. Therefore, in the following graphs, we only show results for S&P 500 stock pool data.

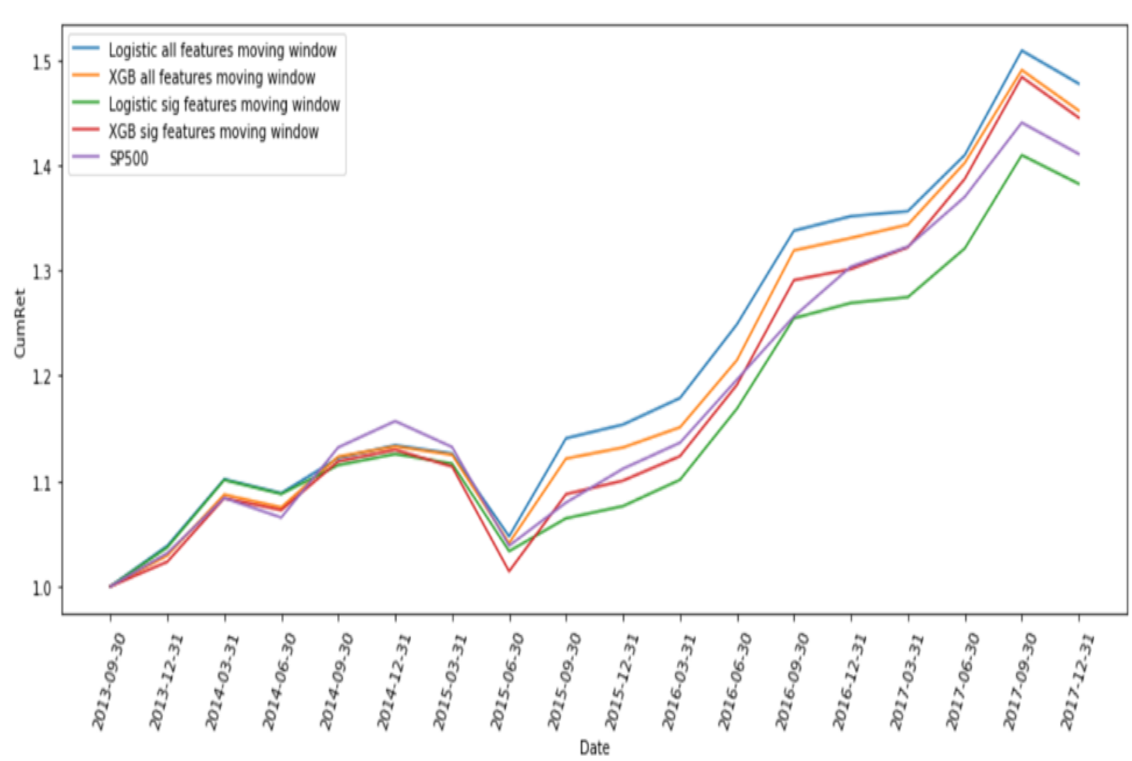


Figure 33: Stock Prediction Model Moving Window Strategy Backtest Result

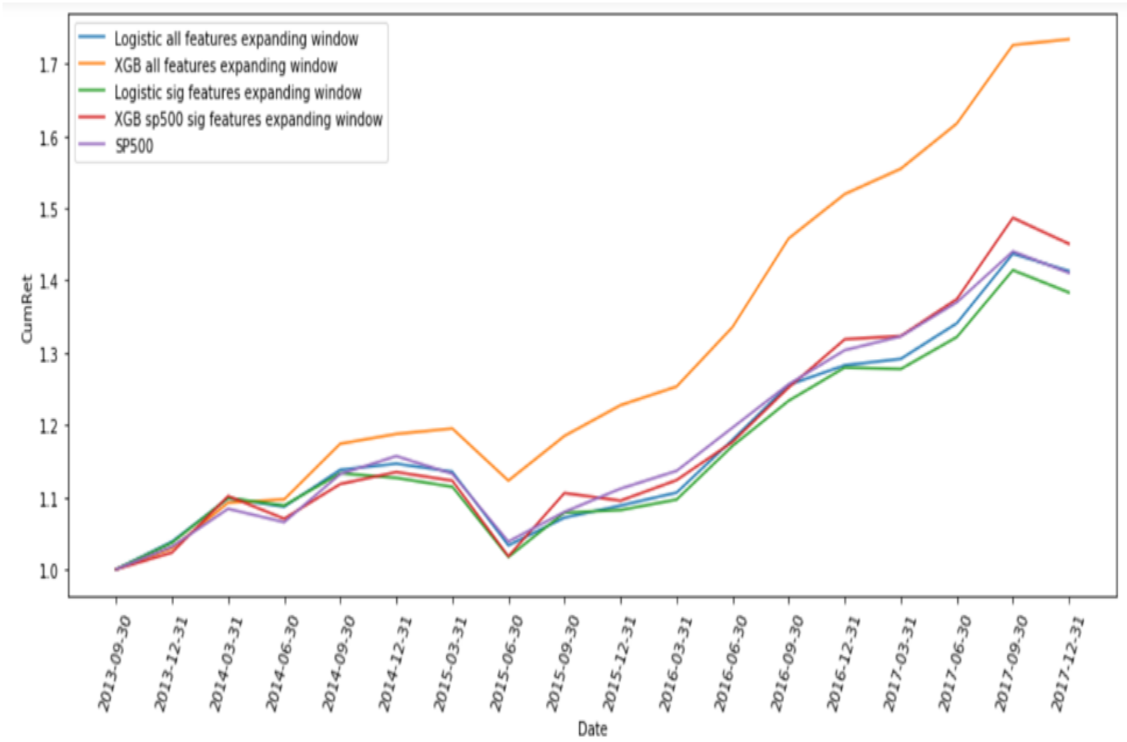


Figure 34: Stock Prediction Model Expanding Window Strategy Backtest Result

We could see that among moving window models, logistic regression with all features included achieved the highest return and 4 out of 5 strategies beat the SP500 benchmark. Among expanding window models however, the XGBoost model with all features included achieved the highest returns. In our moving window models, because the data available was very limited, data-heavy machine learning models such as XGBoost under-performed the simple logistic model. However, in our expanding window's setting, where data was sufficient, XGBoost outperforms other models significantly. It achieved an annualized return of 15 % and a Sharpe Ratio of 0.987.

The result here that expanding window models outperform moving window models is consistent with the results from Copycat strategy that the longer the look back period, the more significant the correlation between past return and next quarterly return.

Moreover, we could observe that for XGBoost models, models with all features significantly outperform models with only significant features. There are two possible explanations. First, the significant features were chosen from the individual correlations with the next quarterly return. Although this is an effective measurement in linear models, it is not necessarily effective for tree-structured models which capture not only the linear relationship between features and dependent variable but also the interaction effect among features. In other word, it is likely that predicting power is low for one individual feature, but increases by considering interaction of two features. Second, by default, the tree structure of XGBoost is limited to a max depth of 6 instead of a full tree to control for over-fitting. Therefore, if a feature is so noisy that it only worsen the model, the tree structure will not consider splitting in that dimension.

The graph below shows the built-in feature importance of our best XGBoost model.

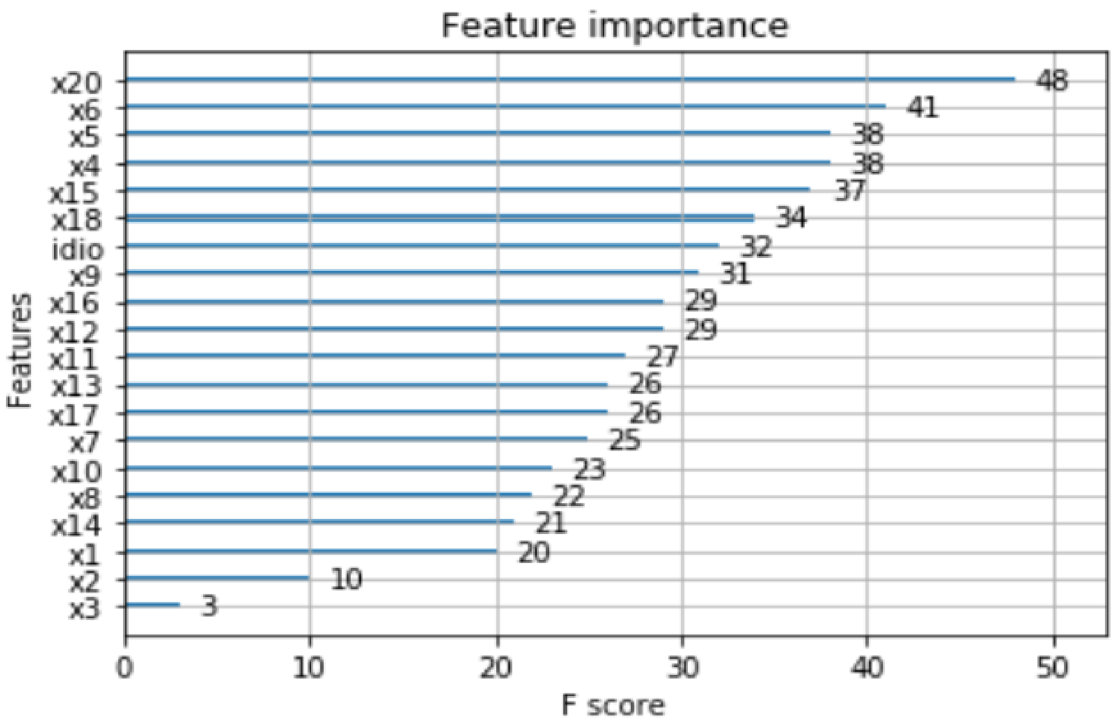


Figure 35: Feature Importance from best XGBoost Model

The results from scatter plots showed that $x4, x6, x8, x10, x11, x16, x17, x18, x19, x20$ were significant features. We can see that they are not necessarily the most significant features in the XGBoost model.

6.3 Copycat Strategy Backtest Results

As previous justification in Section 5, we found that serial return has somewhat positive correlation, which implicitly meant good performance to be consistent. The optimal portfolio should be chosen from past top performed funds. We built our portfolio based on past top 5, 10, 15, 20 performed funds.

The Copycat strategy is to access the holding positions of each funds in a historical period, then replicate and implement their portfolios. We first want to study the Copycat strategy with returns as indicator in this work. For example, we calculated and ranked the funds' returns over the first period, 2013/03/31 to 2013/06/30. Suppose we are only interested in the top 20 performed funds, we copied the stocks holdings of the funds with top 20 returns, and executed the replicated portfolio with the same holdings as them over 2013/09/30 to 2013/12/31. Basically, we constructed our equally weighted portfolio based on the past accessible data when looking back one more filling period. Also, considering the fact that funds that hold too many stocks actually tend to weight on a specific section, which may not reflect the stock selection skills of fund managers, we therefore strictly constrained the number of equities assets in portfolio to be less than 200, also being consistent with the investment policy of Rebellion Research. And we leveraged the weights or fractions of equities assets in portfolios of target funds as the stock selection criteria to satisfy the constraints of stocks number in portfolio construction. Thus, we may only consider the top 10 highly weighted stocks if the top 20 funds are of our interest. So if more than one funds all select the same stocks based on our methodology, we assigned more weights on those stocks. Also, to make some comparisons, we also constructed portfolios based on top and last 5, 10, 15, 20 performed funds, and made the plots of profit and loss as below.

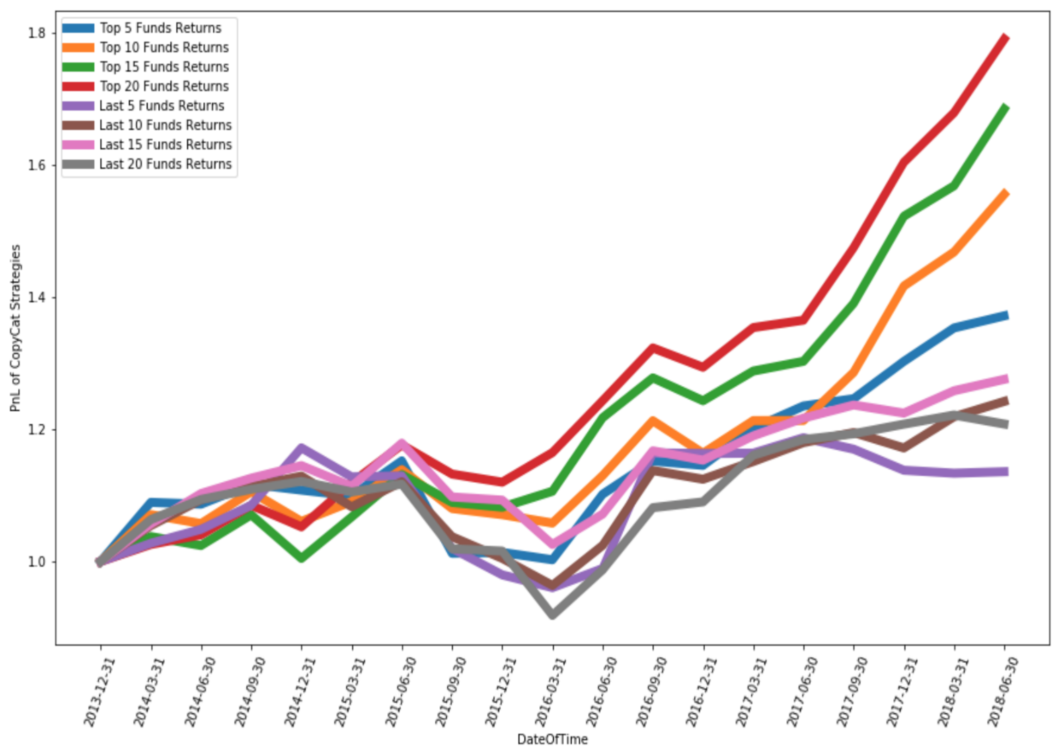


Figure 36: Copycat Strategy Based on Funds Returns

In the results, the Copycat strategy using top 20 funds performs the best. The holding period return of this strategy over our investment horizon is around 79.98%, that is the annualized return is 13.04%. In addition, strategies using less top funds performs less well. We believe one of the reasons is that fund managers who performed well in the past may weight much on similar group stocks. And those stocks were also selected based on the simple stock selection and asset allocation methods. Similarly, we carried out the Copycat strategy based on cash inflow of funds.

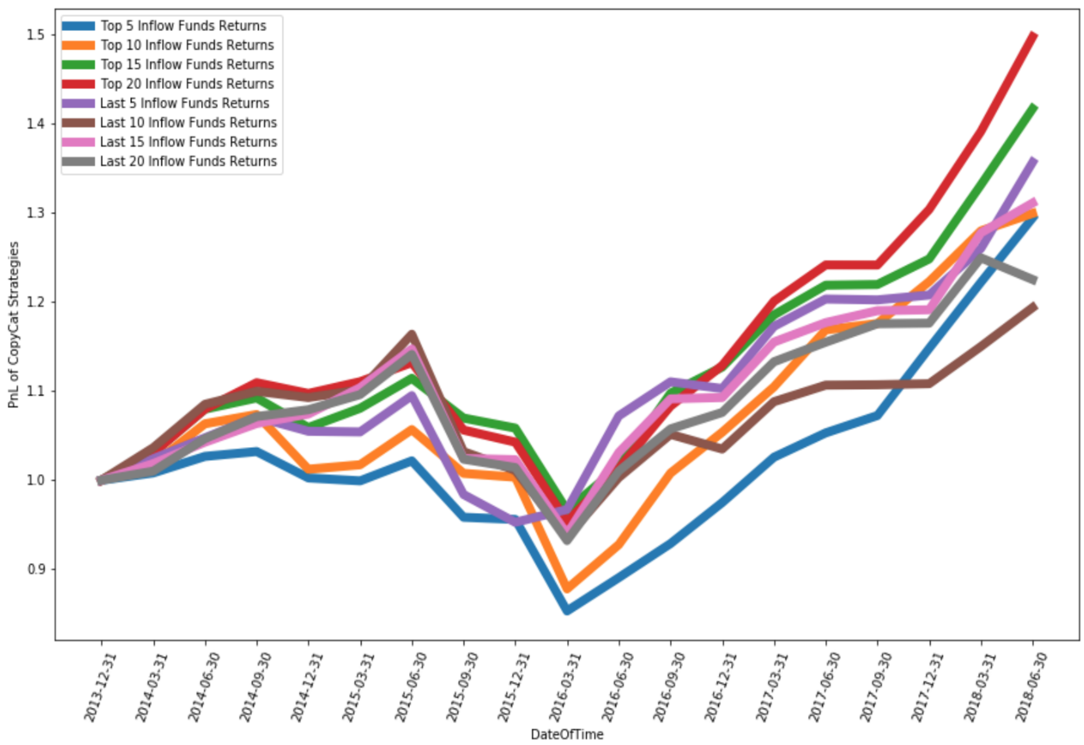


Figure 37: Copycat Strategy Based on Funds Cash Inflows

Based on the plots above, we concluded that the Copycat strategy using 20 funds with top ranked cash inflows performs the best. And this result is consistent with Copycat strategy using returns metrics. In further analyses and comparisons with stock predictions model, we considered strategy using funds with top 20 returns as the best performed.

Another natural approach is to practice the copycat strategy with more than one looking back periods. Recall in the section 5.2, the negative correlation between next quarter returns and historical cumulative returns in past 8 quarters is slightly larger than 1 quarter return. Therefore, we would like to explore the copycat strategies with 8 quarters looking back periods based on top 20 returns.

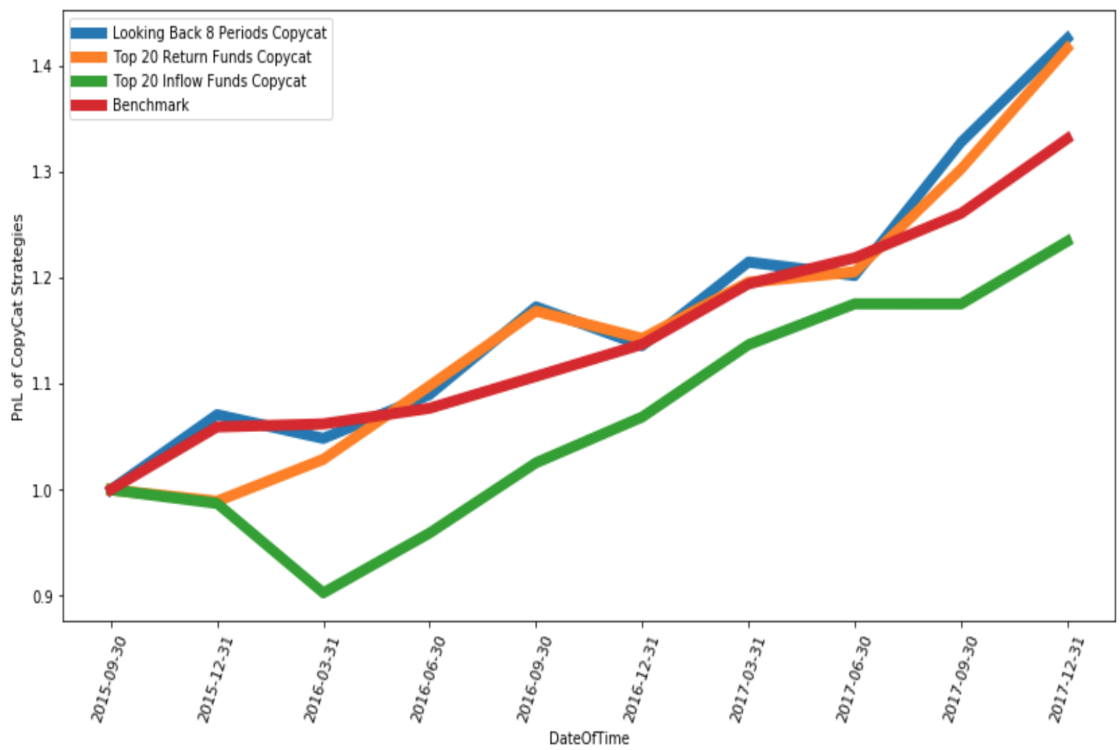


Figure 38: Copycat Strategy Looking Back More Quarters

In this plot, we displayed the three strategies performance and benchmark for comparisons. Notice we need to look back 8 quarters in this case, hence the starting point of our portfolio is 2015/09/30 and we have 2 years for backtest. Before the middle of 2016 and after the end of 2016, the copycat strategy based on 8 quarters looking back and top 20 returns performed slightly better than the strategy based on 1 quarter looking back and top 20 returns, which is close to our expectation and correlation analysis result. The top 20 cash inflow funds copycat strategy under-performed than the S&P500 index benchmark.

Overall, in the exploration of Copycat strategy, we also investigated other approaches such as turnover metrics, sector classification and mean-reverting. However, after the correlation analyses, business intuitions and performance analysis, only returns and cash inflow based strategy are introduced in this work. Furthermore, we combined the copycat strategies based on top returns and top cash inflows together by taking the intersections of stocks in two portfolios over each holding period, and generated the final copycat strategy.

6.4 Summary Statistics and Comparison

In this subsection, we compare the best two models from Stock Prediction Strategy and Copycat Strategy. The best portfolio from Stock Prediction is the portfolio predicted by XGBoost Model with S&P 500 stock pool using moving window with all features. The best portfolio from Copycat is the intersection of stock pools selected from Top 20 Return Copycat and Top 20 Inflow Copycat. Lists of specific stock CUSIPs are listed in the appendix. We both look at the plot of cumulative return and some summary statistics in order to compare the portfolios.

The graph below shows the cumulative returns of our two portfolios from 2013/12/31 to 2017/12/31, together with the benchmark S&P 500.

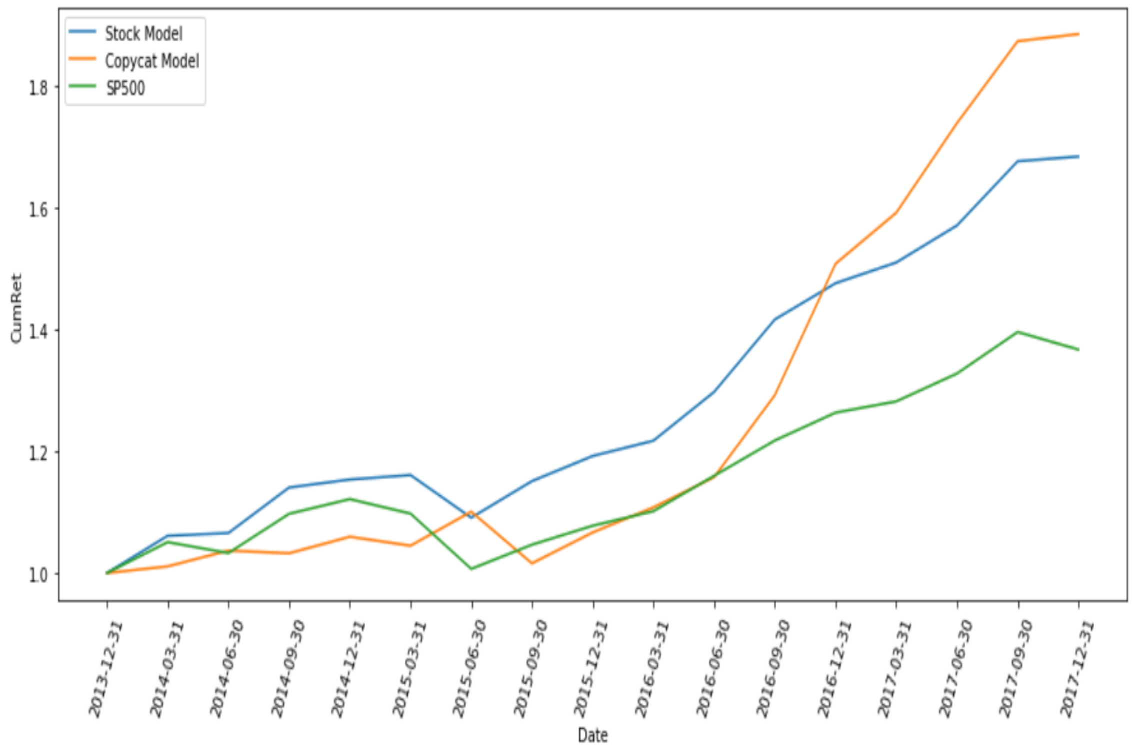


Figure 39: Best Strategies PnL Comparison with S&P 500

The 2 tables below summarize some statistics about the performance and turnover of our strategies. From performance table, we could see that Copycat Strategy achieved the highest annualized return (19.8%) while the Stock Prediction Model achieved the highest Sharpe Ratio (0.987). Both strategies beat the SP500 benchmark.

From the turnover table, we could see that the overall turnover of our strategy is quite high, which may result in huge transaction costs. The average quarterly turnover ratio for the stock prediction model is 0.535 and the average turnover for the copycat strategy is 0.627. Therefore, structuring our trades to reduce turnover and transaction costs is necessary when these strategies are implemented in live trading.

Table 6: Summary Statistics of best performing strategies

Metric	Stock Prediction Model	Copycat Strategy	SP500 Benchmark
Annualized Return	0.150	0.198	0.095
Standard Deviation	0.150	0.255	0.143
Sharpe Ratio	0.987	0.776	0.66
Maximum Drawdown	0.07	0.085	0.118
Calmar Ratio	2.14	2.33	0.803
Average Turnover	0.535	0.627	NA

Table 7: Strategies' Turnover

Time	Stock Prediction Model	Copycat Strategy
2014-03-31	0.59	0.767
2014-06-30	0.46	0.595
2014-09-30	0.555	0.406
2014-12-31'	0.55	0.156
2015-03-31	0.425	0.833
2015-06-30	0.6	0.719
2015-09-30	0.108	0.696
2015-12-31	0.821	0.422
2016-03-31	0.769	0.798
2016-06-30	0.61	0.765
2016-09-30	0.44	0.133
2016-12-31	0.676	1.0
2017-03-31	0.515	0.605
2017-06-30	0.47	0.086
2017-09-30	0.46	0.875
2017-12-31	0.515	0.823

6.5 Limitations and Suggestions

One limitation of our research is that we used only five years data from 2013 to 2017 to construct and test our strategies. Since we have too few data points on time series, it was really difficult for us to apply any complicated time series models. In addition, since we only have recent five years' data, which is just slightly longer than one economic cycle, and the market as a whole has been expanding in the past five years, it is hard to say whether our strategies would perform as good when the economy is in a recession. If we could have historical 13F data in the past twenty years, we would probably be able to construct more solid time series models rather than just to use one quarter cross sectional institutional filings to extract features. Also, data of longer time period is essential for us to modify and validate our strategies to make sure that it is still capable of outperforming the benchmark even in bad times in economy.

One suggestion on expanding current research is to look at the amendment filings. In our research we ignored amendments so as to reduce the workload on data processing, which already took us quite a huge amount of time to clean the raw data. But we still value the holdings data disclosed in amendments because generally, institutions do not want to disclose their most profitable trades and holdings too early to avoid free ride problems. So the institutions tend to request for confidential treatments from SEC and publish their secret holdings later. In addition to this, funds also use amendment filings to correct misstated positions in their original filings. So the original securities positions data could have some mistakes. Thus, extracting information in the amendments could be our next steps to improve our current results. We could devise a model to focus on those data to see how much portion of the total holdings are disclosed later and how do those securities perform.

Another suggestion for further research is that we could include more fund level features. We computed the three most important features to measure fund performance, but other fund level metrics could also be important trading signals, such as the securities bought and sold by funds with better historical performance, and the years of survival should we have longer period data. Furthermore, as for stock level prediction, we used only XGBoost and logistic regression, there are many more machine learning models to explore and apply to our 13F data, such as SVM, CNN, etc.

7 Appendix

7.1 List of stocks - Stock Prediction Strategy

2013/12/31: ['002824100', '00724F101','00817Y108', '009158106', '00971T101', '011659109', '015351109', '018490102', '021441100', '02209S103', '023135106', '02503Y103', '025537101', '025816109', '03073E105', '03076C106', '03748R101', '037604105', '037833100', '039483102', '052769106', '053015103', '053611109', '054937107', '064058100', '071813109', '075896100', '097023105', '110122108', '111320107', '125509109', '12572Q105', '126408103', '126650100', '12673P105', '14040H105', '14149Y108', '143658300', '149123101', '166764100', '171232101', '171340102', '17275R102', '172967424', '177376100', '189054109', '191216100', '192446102', '194162103', '200340107', '205887102', '21036P108', '219350105', '22160K105', '23331A109', '235851102', '237194105', '23918K108', '244199105', '247361702', '254687106', '254709108', '25470M109', '25490A309', '256677105', '260003108', '26138E109', '263534109', '26441C204', '278865100', '28176E108', '30212P303', '30225T102', '30231G102', '313586109', '314211103', '316773100', '320517105', '339041105', '343412102', '34354P105', '344849104', '34964C106', '35671D857', '369550108', '369604103', '370334104', '37045V100', '372460105', '37247D106', '375558103', '416515104', '418056107', '436106108', '437076102', '438516106', '444859102', '44919P508', '452308109', '452327109', '458140100', '460146103', '461202103', '46625H100', '466313103', '478160104', '482480100', '494368103', '501797104', '524660107', '524901105', '526057104', '532457108', '534187109', '539830109', '55261F104', '55616P104', '565849106', '56585A102', '571748102', '571903202', '577081102', '580135101', '58155Q103', '58933Y105', '59156R108', '594918104', '595017104', '595112103', '608190104', '60871R209', '609207105', '61166W101', '611742107', '615369105', '617446448', '620076307', '640268108', '64110L106', '654106103', '655844108', '665859104', '666807102', '681919106', '682680103', '68389X105', '693475105', '693506107', '693656100', '712704105', '713448108', '717081103', '718546104', '74144T108', '74251V102', '742718109', '744320102', '745867101', '747525103', '74762E102', '755111507', '756577102', '75886F107', '7591EP100', '79466L302', '808513105', '816851109', '83088M102', '831865209', '844741108', '857477103', '863667101', '871829107', '872540109', '882508104', '883556102', '88579Y101', '88732J207', '892356106', '90130A101', '904311107', '910047109', '911312106', '911363109', '913017109', '91529Y106', '92343V104', '92826C839', '931142103', '949746101', '959802109', '963320106', '988498101', '989701107', 'G0408V102', 'G1151C101', 'G2554F113', 'G29183103', 'G60754101', 'G7945M107']

2014/03/31: ['002824100', '002896207', '007903107', '00817Y108', '018490102', '020002101', '02076X102', '02209S103', '025537101', '025816109', '026874784', '032511107', '032654105', '038222105', '039483102', '053015103', '053774105', '054937107', '057224107', '058498106', '060505104', '064058100', '110122108', '111320107', '124857202', '12504L109', '125269100', '125896100', '126408103', '126650100', '12686C109', '127097103', '127387108', '13342B105', '14040H105', '14149Y108', '143130102', '149123101', '156700106', '177376100', '192446102', '194162103', '20030N101', '20825C104', '219350105', '235851102', '247361702', '25179M103', '254709108', '25470M109', '25490A309', '256746108', '25746U109', '26138E109', '263534109', '268648102', '26875P101', '269246401', '278642103', '281020107', '28176E108', '30161N101', '30249U101', '30303M102', '311900104', '313586109', '316773100', '320517105', '337738108', '344849104', '364760108', '369550108', '370334104', '37247D106', '382550101', '406216101', '410345102', '416515104', '42809H107', '436106108', '436440101', '437076102', '44107P104', '443683107', '444859102', '446150104', '452308109', '460690100', '487836108', '493267108', '49446R109', '49456B101', '500255104', '50076Q106', '501044101', '501797104', '512807108', '532457108', '534187109', '535678106', '539830109', '544147101', '548661107', '55262C100', '552848103', '552953101', '55616P104', '565849106', '56585A102', '571748102', '571903202', '574599106', '580135101', '582839106', '58933Y105', '59156R108', '595017104', '595112103', '609207105', '61166W101', '611742107', '617446448', '61945C103', '629377508', '64110D104', '651290108', '651639106', '65339F101', '654106103', '65473P105', '655044105', '655844108', '670346105', '67066G104', '674599105', '676220106', '68389X105', '69331C108', '693475105', '693506107', '69351T106', '693718108', '708160106', '716768106', '718546104', '724479100', '744573106', '74733V100', '747525103', '74834L100', '755111507', '7591EP100', '760759100', '761713106', '778296103', '786514208', '790849103', '80004C101', '806857108', '808513105', '812350106', '83088M102', '842587107', '845467109', '867914103', '871829107', '872540109', '874054109', '87612E106', '88033G407', '88076W103', '880779103', '882508104', '883203101', '887317303', '896945201', '902494103', '902973304', '904311107', '907818108', '913017109', '91324P102', '918204108', '91913Y100', '92826C839', '958102105', '959802109', '969457100', '98212B103', '98389B100', '988498101', '989701107', 'G2554F113', 'G29183103', 'G491BT108', 'G6359F103', 'H84989104', 'H8817H100', 'N53745100', 'V7780T103']

2014/06/30:['00130H105', '002824100', '002896207', '00507V109', '00724F101', '00846U101', '00971T101', '018490102', '020002101', '02209S103', '023135106', '025537101', '026874784', '03027X100', '031162100', '032511107', '037411105', '038222105', '053015103', '053774105', '057224107', '058498106', '060505104', '064058100', '075896100', '086516101', '093671105', '097023105', '111320107', '124857202', '12504L109', '12541W209', '125896100', '12673P105', '13342B105', '14040H105', '14149Y108', '143130102', '149123101', '151020104', '165167107', '177376100', '194162103', '20030N101', '20825C104', '219350105', '22160K105', '247361702', '247916208', '25179M103', '253868103', '254709108', '256677105', '26138E109', '264411505', '26875P101', '278642103', '285512109', '29364G103', '29476L107', '30161N101', '30212P303', '30303M102', '307000109', '31428X106', '337738108', '337932107', '345370860', '34959E109', '354613101', '35671D857', '364760108', '369550108', '370334104', '37045V100', '375558103', '37940X102', '38141G104', '382550101', '40412C101', '406216101', '410345102', '416515104', '42809H107', '436106108', '436440101', '438516106', '44107P104', '452308109', '45337C102', '459902102', '460146103', '460690100', '46120E602', '466313103', '47102X105', '485865109', '48666K109', '49271M100', '493267108', '49446R109', '49456B101', '50076Q106', '501044101', '526057104', '532457108', '534187109', '544147101', '546347105', '552848103', '552953101', '55616P104', '565849106', '571748102', '571903202', '574599106', '57772K101', '58933Y105', '59156R108', '595112103', '609207105', '61166W101', '611742107', '617446448', '61945C103', '629377508', '631103108', '637071101', '651229106', '651290108', '65339F101', '65473P105', '655044105', '674599105', '681919106', '69331C108', '693475105', '718172109', '718546104', '743315103', '744320102', '745867101', '74733V100', '756109104', '756577102', '7591EP100', '761713106', '778296103', '790849103', '80004C101', '806857108', '81211K100', '83088M102', '844741108', '847560109', '855244109', '85590A401', '857477103', '863667101', '867914103', '871829107', '88033G407', '882508104', '883556102', '887317303', '88732J207', '89417E109', '896945201', '902494103', '902973304', '904311107', '907818108', '911312106', '911363109', '912909108', '91324P102', '92826C839', '94106L109', '958102105', '959802109', '962166104', '969457100', '98212B103', '988498101', '98956P102', '98978V103', 'G1151C101', 'G2554F113', 'G29183103', 'G3157S106', 'G47791101', 'G491BT108', 'G5785G107', 'G6359F103', 'G65431101', 'G7665A101', 'G7945M107', 'H8817H100', 'N53745100', 'V7780T103']

2014/09/30:['001055102', '00130H105', '002824100', '00287Y109', '00724F101', '007903107', '00817Y108', '00971T101', '018490102', '02076X102', '021441100', '023135106', '023608102', '02503Y103', '025537101', '025816109', '026874784', '03073E105', '032654105', '037833100', '038222105', '039483102', '054303102', '054937107', '058498106', '075896100', '086516101', '093671105', '097023105', '099724106', '101137107', '124857202', '125509109', '12572Q105', '125896100', '126408103', '127097103', '127387108', '14040H105', '14170T101', '143658300', '149123101', '15135B101', '15189T107', '156782104', '165167107', '171340102', '192446102', '20030N101', '209115104', '219350105', '22160K105', '231021106', '23331A109', '235851102', '237194105', '242370203', '247916208', '25179M103', '254709108', '25470M109', '25490A309', '256677105', '25746U109', '264411505', '26441C204', '26875P101', '269246401', '285512109', '29364G103', '30161N101', '30303M102', '31428X106', '316773100', '337932107', '354613101', '35671D857', '36467W109', '369604103', '370334104', '37045V100', '37247D106', '382550101', '40412C101', '40414L109', '406216101', '412822108', '416515104', '42809H107', '438516106', '440452100', '452308109', '466313103', '481165108', '48203R104', '487836108', '49271M100', '493267108', '49446R109', '50076Q106', '501889208', '512807108', '518439104', '526057104', '52729N308', '532457108', '534187109', '535678106', '544147101', '546347105', '552848103', '552953101', '55616P104', '565849106', '571748102', '57636Q104', '577081102', '57772K101', '58933Y105', '59156R108', '595017104', '60871R209', '61166W101', '61945C103', '62886E108', '629377508', '637071101', '651290108', '65339F101', '65473P105', '655044105', '655844108', '670346105', '674599105', '676220106', '681919106', '68389X105', '69331C108', '704326107', '712704105', '716768106', '718172109', '718546104', '743315103', '744573106', '745867101', '747525103', '75281A109', '756109104', '7591EP100', '760759100', '761713106', '790849103', '79466L302', '80004C101', '806857108', '842587107', '845467109', '847560109', '85590A401', '867914103', '871829107', '874054109', '88033G407', '880770102', '880779103', '882508104', '883203101', '883556102', '88579Y101', '88732J207', '89417E109', '896945201', '90130A101', '902973304', '907818108', '910047109', '911363109', '91324P102', '917047102', '92826C839', '949746101', '959802109', '962166104', '969457100', '98212B103', '98389B100', '988498101', 'G1151C101', 'G29183103', 'G47791101', 'G48833100', 'G5785G107', 'G60754101', 'G65431101', 'G7945M107', 'G7S00T104', 'H2906T109', 'H84989104', 'N53745100']

2014/12/31: ['001055102', '00724F101', '00846U101', '020002101', '021441100', '02209S103', '02376R102', '025537101', '025816109', '026874784', '037411105', '038222105', '039483102', '052769106', '053015103', '053774105', '054303102', '054937107', '060505104', '064058100', '071813109', '075896100', '086516101', '093671105', '097023105', '110122108', '124857202', '125896100', '126408103', '126650100', '12673P105', '12686C109', '127387108', '13342B105', '14040H105', '143130102', '143658300', '171779309', '172967424', '192446102', '194162103', '20030N101', '209115104', '219350105', '22160K105', '23331A109', '235851102', '242370203', '244199105', '247361702', '254709108', '25470M109', '256677105', '256746108', '25746U109', '263534109', '26441C204', '268648102', '269246401', '278642103', '281020107', '28176E108', '285512109', '29364G103', '30161N101', '30219G108', '311900104', '31428X106', '316773100', '320517105', '337738108', '337932107', '345370860', '34959E109', '354613101', '35671D857', '364760108', '370334104', '37045V100', '38141G104', '382550101', '40414L109', '406216101', '416515104', '436440101', '438516106', '44107P104', '443683107', '446150104', '452308109', '45866F104', '460146103', '460690100', '466313103', '47102X105', '48203R104', '485865109', '487836108', '493267108', '49446R109', '49456B101', '50076Q106', '501044101', '512807108', '526057104', '52729N308', '532457108', '534187109', '535678106', '544147101', '546347105', '548661107', '552848103', '55616P104', '56585A102', '571748102', '571903202', '574599106', '57636Q104', '57772K101', '580135101', '58933Y105', '59156R108', '595017104', '595112103', '609207105', '61166W101', '617446448', '61945C103', '620076307', '64110D104', '64110L106', '651229106', '651639106', '65339F101', '65473P105', '655044105', '674599105', '676220106', '69331C108', '693475105', '693506107', '69351T106', '693718108', '712704105', '724479100', '743315103', '74340W103', '744320102', '744573106', '745867101', '756109104', '756577102', '7591EP100', '778296103', '79466L302', '808513105', '81211K100', '83088M102', '842587107', '844741108', '855030102', '85590A401', '857477103', '867914103', '872375100', '872540109', '880770102', '883203101', '887317303', '88732J207', '89417E109', '896945201', '90130A101', '90184L102', '902973304', '907818108', '910047109', '911312106', '91324P102', '91913Y100', '92276F100', '92826C839', '931142103', '94106L109', '958102105', '959802109', '962166104', '966837106', '98389B100', '983919101', '989701107', 'G1151C101', 'G29183103', 'G30401106', 'G47791101', 'G7945M107', 'H8817H100', 'N53745100', 'V7780T103']

2015/03/31: ['001055102', '00206R102', '002824100', '00287Y109', '00507V109', '007903107', '00817Y108', '00846U101', '020002101', '023135106', '025537101', '025816109', '026874784', '03073E105', '031162100', '032511107', '032654105', '036752103', '037833100', '038222105', '053015103', '054937107', '057224107', '058498106', '064058100', '071813109', '09062X103', '097023105', '099724106', '101137107', '110122108', '111320107', '124857202', '12504L109', '12541W209', '125896100', '126408103', '126650100', '12673P105', '14040H105', '149123101', '165167107', '172967424', '192446102', '194162103', '20605P101', '20825C104', '209115104', '219350105', '22160K105', '23331A109', '235851102', '247916208', '254687106', '25490A309', '256677105', '256746108', '25746U109', '260003108', '26441C204', '269246401', '278642103', '281020107', '28176E108', '285512109', '30161N101', '30212P303', '30219G108', '30303M102', '311900104', '316773100', '336433107', '337738108', '344849104', '345370860', '34959E109', '354613101', '36467W109', '37045V100', '38141G104', '382550101', '40412C101', '40414L109', '406216101', '410345102', '416515104', '423452101', '436106108', '436440101', '437076102', '438516106', '441060100', '443683107', '446150104', '452308109', '458140100', '45866F104', '460146103', '46625H100', '466313103', '485865109', '487836108', '493267108', '500255104', '501044101', '518439104', '526057104', '52729N308', '534187109', '548661107', '552848103', '552953101', '56585A102', '571748102', '574599106', '57772K101', '58933Y105', '59156R108', '611742107', '617446448', '61945C103', '637071101', '651290108', '651639106', '65339F101', '654106103', '655044105', '655844108', '670346105', '676220106', '682680103', '68389X105', '69331C108', '693475105', '704326107', '712704105', '717081103', '718172109', '718546104', '723787107', '742718109', '743315103', '745867101', '74762E102', '75281A109', '778296103', '790849103', '79466L302', '806857108', '808513105', '83088M102', '842587107', '844741108', '847560109', '855244109', '85590A401', '857477103', '867914103', '871503108', '871829107', '872540109', '87612E106', '880779103', '883203101', '883556102', '88579Y101', '887317303', '88732J207', '89417E109', '90184L102', '902973304', '904311107', '907818108', '911312106', '911363109', '913017109', '91324P102', '917047102', '91913Y100', '92276F100', '92343V104', '92826C839', '931142103', '931427108', '94106L109', '949746101', '959802109', '962166104', '969457100', '988498101', '989701107', '98978V103', 'G1151C101', 'G29183103', 'G30401106', 'G47791101', 'G48833100', 'G491BT108', 'G6359F103', 'N53745100']

2015/06/30:['00130H105', '00206R102', '002824100', '00287Y109', '00817Y108', '021441100', '023135106', '023608102', '02376R102', '025537101', '025816109', '026874784', '03027X100', '03073E105', '031162100', '032654105', '036752103', '039483102', '053015103', '054937107', '057224107', '060505104', '064058100', '09062X103', '097023105', '099724106', '110122108', '115637209', '124857202', '125269100', '125509109', '126650100', '127387108', '13342B105', '134429109', '14040H105', '143658300', '149123101', '151020104', '15135B101', '171779309', '17275R102', '172967424', '192446102', '20030N101', '205887102', '209115104', '219350105', '22822V101', '231021106', '23331A109', '235851102', '25179M103', '254687106', '254709108', '25470F104', '25470M109', '256677105', '256746108', '25746U109', '260003108', '26441C204', '268648102', '269246401', '277432100', '278642103', '28176E108', '291011104', '29364G103', '30161N101', '30212P303', '30219G108', '30303M102', '311900104', '316773100', '320517105', '337738108', '343412102', '344849104', '34959E109', '35671D857', '36467W109', '370334104', '37045V100', '375558103', '38141G104', '382550101', '40412C101', '406216101', '410345102', '418056107', '436106108', '436440101', '437076102', '438516106', '44107P104', '444859102', '446150104', '452308109', '452327109', '46625H100', '469814107', '482480100', '487836108', '49271M100', '493267108', '494368103', '49456B101', '501889208', '518439104', '532457108', '534187109', '539830109', '552848103', '55616P104', '56585A102', '571748102', '57636Q104', '58933Y105', '59156R108', '595017104', '609207105', '617446448', '62886E108', '640268108', '651229106', '651639106', '65339F101', '654106103', '65473P105', '665859104', '681919106', '69331C108', '693475105', '693506107', '693718108', '695156109', '704326107', '712704105', '718172109', '740189105', '74251V102', '743315103', '744320102', '745867101', '74762E102', '7591EP100', '760759100', '761713106', '790849103', '808513105', '83088M102', '842587107', '844741108', '845467109', '847560109', '855244109', '857477103', '867914103', '871503108', '874054109', '87612E106', '88033G407', '880770102', '882508104', '883203101', '88579Y101', '886547108', '887317303', '88732J207', '90184L102', '902494103', '902973304', '904311107', '911312106', '913017109', '91324P102', '91529Y106', '92343V104', '92532F100', '92826C839', '931427108', '949746101', '959802109', '98212B103', '983919101', '989701107', '98978V103', 'G1151C101', 'G29183103', 'G3157S106', 'G47791101', 'G491BT108', 'G5960L103', 'G60754101', 'G66721104', 'G97822103', 'N53745100', 'N59465109', 'V7780T103']

2015/09/30: ['001055102', '003654100', '00507V109', '00724F101', '007903107', '00971T101', '011659109', '012653101', '015271109', '018581108', '018802108', '02376R102', '030420103', '032654105', '05329W102', '053332102', '081437105', '097023105', '099724106', '101121101', '101137107', '115637209', '12503M108', '12504L109', '125509109', '12572Q105', '127097103', '143130102', '15135B101', '205887102', '21036P108', '217204106', '22160K105', '247916208', '25271C102', '253868103', '254067101', '25470M109', '256746108', '260003108', '264411505', '26441C204', '26483E100', '26875P101', '26884L109', '269246401', '278642103', '28176E108', '285512109', '294429105', '29476L107', '297178105', '30219G108', '302445101', '30249U101', '311900104', '320517105', '337932107', '34354P105', '34988V106', '36467W109', '369550108', '372460105', '410345102', '412822108', '413086109', '413875105', '423452101', '42809H107', '446150104', '45168D104', '46284V101', '482480100', '49271M100', '500255104', '535678106', '55261F104', '552848103', '56585A102', '595017104', '615369105', '61945C103', '620076307', '62886E108', '63938C108', '651290108', '65249B208', '65339F101', '655664100', '676220106', '681919106', '693506107', '712704105', '714046109', '718546104', '723787107', '729251108', '744573106', '74733V100', '751212101', '75281A109', '756109104', '758849103', '7591EP100', '761713106', '774341101', '776696106', '78440X101', '806407102', '808513105', '824348106', '83088M102', '844741108', '845467109', '847560109', '854502101', '855030102', '855244109', '871503108', '871607107', '87165B103', '874054109', '887317303', '892356106', '902653104', '90384S303', '917047102', '91913Y100', '92276F100', '92345Y106', '92826C839', '941848103', '966837106', 'G47791101', 'G5785G107', 'G6359F103', 'G65431101', 'G7S00T104', 'G91442106', 'H2906T109', 'N59465109']

2015/12/31: ['001055102', '00724F101', '012653101', '015271109', '02079K305', '02209S103', '023135106', '026874784', '03027X100', '031100100', '03662Q105', '036752103', '037411105', '040413106', '04621X108', '053015103', '053611109', '064058100', '086516101', '09247X101', '097023105', '115637209', '124857202', '12503M108', '12572Q105', '127387108', '149123101', '172908105', '172967424', '189054109', '194162103', '20030N101', '21036P108', '22160K105', '231021106', '23918K108', '25271C102', '254709108', '260003108', '29476L107', '31428X106', '337738108', '343412102', '34964C106', '354613101', '370334104', '37045V100', '38141G104', '382550101', '436440101', '437076102', '438516106', '445658107', '446150104', '452308109', '466313103', '469814107', '50540R409', '518439104', '534187109', '535678106', '546347105', '552953101', '554382101', '56585A102', '574599106', '579780206', '59156R108', '595017104', '59522J103', '609207105', '615369105', '617446448', '64110L106', '654106103', '655664100', '693475105', '693506107', '69351T106', '712704105', '718172109', '724479100', '74144T108', '744320102', '761713106', '770323103', '78440X101', '80589M102', '844741108', '847560109', '854502101', '855244109', '863667101', '871829107', '87612E106', '88579Y101', '88732J207', '893641100', '904311107', '907818108', '910047109', '911312106', '911363109', '91529Y106', '918204108', '91913Y100', '92826C839', '929160109', '94106L109', '941848103', '983919101', 'G0177J108', 'G1151C101', 'G29183103', 'G47791101', 'G491BT108', 'G6518L108', 'H84989104']

2016/03/31: ['00724F101', '00751Y106', '00817Y108', '011659109', '02079K305', '025537101', '03027X100', '039483102', '053484101', '093671105', '15189T107', '177376100', '209115104', '23331A109', '235851102', '237194105', '23918K108', '253868103', '25470F104', '256677105', '256746108', '278865100', '281020107', '29364G103', '294429105', '30040W108', '30225T102', '314211103', '31620M106', '34964C106', '40412C101', '40434L105', '406216101', '412822108', '436106108', '469814107', '47102X105', '512807108', '526057104', '52729N308', '57636Q104', '57772K101', '59522J103', '608190104', '60871R209', '620076307', '629377508', '64110D104', '651229106', '654106103', '666807102', '67066G104', '69351T106', '712704105', '718172109', '723787107', '74460D109', '774341101', '80589M102', '811065101', '83088M102', '842587107', '844741108', '855244109', '858912108', '871829107', '874054109', '902494103', '90384S303', '904311107', '913903100', '92276F100', '92345Y106', '94946T106', 'G0176J109', 'G1151C101', 'G3223R108', 'G6359F103']

2016/06/30: ['002824100', '00287Y109', '00507V109', '007903107', '00817Y108', '00846U101', '00971T101', '011659109', '012653101', '02079K305', '023135106', '02503Y103', '025816109', '03027X100', '031162100', '032511107', '032654105', '037411105', '037833100', '038222105', '039483102', '053015103', '054303102', '054937107', '057224107', '060505104', '064058100', '071813109', '075896100', '093671105', '099724106', '101137107', '110122108', '126408103', '126650100', '12673P105', '14149Y108', '149123101', '151020104', '15189T107', '156700106', '156782104', '166764100', '171340102', '171779309', '17275R102', '172967424', '20030N101', '219350105', '22822V101', '231021106', '23331A109', '235851102', '237194105', '244199105', '247361702', '25179M103', '25271C102', '253868103', '254709108', '25470M109', '256677105', '256746108', '263534109', '268648102', '26875P101', '26884L109', '278642103', '285512109', '291011104', '29476L107', '30219G108', '31620M106', '316773100', '336433107', '337738108', '344849104', '34959E109', '34964C106', '363576109', '364760108', '375558103', '40434L105', '406216101', '416515104', '423452101', '42809H107', '438516106', '440452100', '458140100', '45866F104', '460146103', '460690100', '46284V101', '46625H100', '466313103', '48203R104', '48666K109', '493267108', '501044101', '501889208', '512807108', '524901105', '526057104', '52729N308', '534187109', '535678106', '548661107', '552953101', '571748102', '577081102', '58155Q103', '58933Y105', '59156R108', '595017104', '609207105', '61166W101', '617446448', '61945C103', '637071101', '64110D104', '64110L106', '651229106', '651290108', '654106103', '655044105', '655664100', '670346105', '67066G104', '674599105', '682680103', '690768403', '69331C108', '704326107', '712704105', '718546104', '723787107', '74251V102', '743315103', '74340W103', '745867101', '74733V100', '747525103', '74762E102', '75281A109', '755111507', '7591EP100', '778296103', '790849103', '79466L302', '806857108', '808513105', '83088M102', '844741108', '847560109', '855244109', '867914103', '872540109', '87612E106', '880770102', '882508104', '883203101', '88579Y101', '886547108', '887317303', '896945201', '902653104', '902973304', '907818108', '911363109', '91324P102', '917047102', '91913Y100', '92532F100', '925550105', '92826C839', '931142103', '94106L109', '958102105', '959802109', '96145D105', '966837106', '983134107', '983919101', '988498101', '98956P102', '98978V103', 'G1151C101', 'G29183103', 'G47791101', 'G5960L103', 'G6359F103', 'G66721104', 'G7665A101', 'G7945M107', 'G81276100', 'G97822103', 'H84989104', 'H8817H100', 'N59465109']

2016/09/30: ['00206R102', '002824100', '00287Y109', '00724F101', '007903107', '00846U101', '009158106', '00971T101', '011659109', '01741R102', '02209S103', '023135106', '02503Y103', '025816109', '030420103', '031100100', '031162100', '032095101', '038222105', '039483102', '052769106', '054937107', '058498106', '064058100', '071813109', '075896100', '086516101', '09062X103', '097023105', '099724106', '101137107', '124857202', '12504L109', '12572Q105', '125896100', '126408103', '14040H105', '143130102', '143658300', '149123101', '151020104', '156782104', '16119P108', '165167107', '171779309', '17275R102', '177376100', '194162103', '205887102', '20605P101', '209115104', '217204106', '219350105', '231021106', '23331A109', '25179M103', '254709108', '25470F104', '260003108', '26138E109', '263534109', '26441C204', '26875P101', '26884L109', '269246401', '278642103', '28176E108', '285512109', '30040W108', '30161N101', '30219G108', '30249U101', '30303M102', '31428X106', '336433107', '337932107', '34354P105', '344849104', '345370860', '34959E109', '354613101', '363576109', '36467W109', '369550108', '37045V100', '37940X102', '38141G104', '382550101', '40434L105', '42809H107', '436106108', '440452100', '444859102', '446150104', '452308109', '452327109', '45866F104', '460146103', '466313103', '485170302', '493267108', '494368103', '49456B101', '500255104', '501044101', '501889208', '512807108', '524901105', '534187109', '548661107', '55616P104', '56585A102', '574599106', '57636Q104', '577081102', '59156R108', '595017104', '595112103', '609207105', '61166W101', '617446448', '637071101', '63938C108', '64110D104', '651229106', '65339F101', '655844108', '665859104', '67066G104', '681919106', '693475105', '693718108', '712704105', '717081103', '718172109', '718546104', '723787107', '74144T108', '744320102', '744573106', '74733V100', '747525103', '74834L100', '75281A109', '755111507', '756577102', '7591EP100', '760759100', '79466L302', '808513105', '83088M102', '831865209', '832696405', '844741108', '855244109', '857477103', '867914103', '871503108', '87165B103', '874054109', '880770102', '880779103', '883203101', '883556102', '88579Y101', '886547108', '887317303', '896945201', '90184L102', '902494103', '902973304', '907818108', '910047109', '911312106', '911363109', '913017109', '91529Y106', '91913Y100', '92343V104', '92826C839', '92939U106', '959802109', '96145D105', '962166104', '983134107', '983919101', '98419M100', '988498101', '98956P102', 'G25839104', 'G29183103', 'G47791101', 'G491BT108', 'G65431101', 'G7945M107', 'G7S00T104', 'H2906T109', 'H84989104', 'H8817H100', 'N53745100']

2016/12/31: ['00206R102', '002824100', '00287Y109', '00507V109', '00724F101', '007903107', '01741R102', '020002101', '02209S103', '023135106', '02376R102', '025816109', '026874784', '031100100', '032511107', '032654105', '037411105', '038222105', '054937107', '057224107', '060505104', '064058100', '071813109', '086516101', '097023105', '099724106', '101137107', '110122108', '124857202', '12504L109', '125269100', '126408103', '126650100', '14040H105', '143130102', '143658300', '149123101', '15189T107', '156700106', '156782104', '165167107', '166764100', '171779309', '172967424', '174610105', '192446102', '194162103', '20030N101', '200340107', '205887102', '20825C104', '219350105', '22160K105', '22822V101', '244199105', '247361702', '247916208', '25179M103', '25271C102', '254687106', '254709108', '25470F104', '256746108', '25746U109', '263534109', '264411505', '26441C204', '26875P101', '269246401', '28176E108', '291011104', '29476L107', '30231G102', '30303M102', '311900104', '313400301', '316773100', '337738108', '345370860', '35671D857', '369604103', '370334104', '37045V100', '38141G104', '40414L109', '40434L105', '406216101', '412822108', '416515104', '42809H107', '42824C109', '436106108', '437076102', '438516106', '440452100', '44107P104', '446150104', '45866F104', '459200101', '460146103', '460690100', '46625H100', '48203R104', '482480100', '493267108', '501044101', '526057104', '52729N308', '532457108', '548661107', '552848103', '552953101', '55616P104', '565849106', '56585A102', '571748102', '571903202', '580135101', '59156R108', '594918104', '595112103', '617446448', '61945C103', '626717102', '629377508', '637071101', '63938C108', '64110D104', '651229106', '651290108', '65339F101', '654106103', '65473P105', '655044105', '670346105', '67066G104', '676220106', '68389X105', '693475105', '70450Y103', '708160106', '712704105', '718172109', '743315103', '74340W103', '744320102', '744573106', '74733V100', '7591EP100', '761713106', '778296103', '79466L302', '806857108', '808513105', '844741108', '855244109', '867914103', '87165B103', '871829107', '872540109', '88033G407', '882508104', '887317303', '90130A101', '902494103', '902973304', '907818108', '910047109', '911312106', '912909108', '913017109', '91324P102', '91913Y100', '92276F100', '92343V104', '92826C839', '931427108', '949746101', '958102105', '959802109', '962166104', '966837106', '98212B103', '983919101', '988498101', '989701107', '98978V103', 'G0177J108', 'G1151C101', 'G29183103', 'G3157S106', 'G47567105', 'G491BT108', 'G5960L103', 'G6359F103', 'G65431101', 'G7665A101', 'G7945M107', 'H8817H100', 'N53745100']

2017/03/31: ['002824100', '00287Y109', '00507V109', '00724F101', '007903107', '00817Y108', '00846U101', '015351109', '020002101', '02209S103', '023135106', '02376R102', '025537101', '025816109', '026874784', '03027X100', '031162100', '032511107', '032654105', '037411105', '038222105', '052769106', '054303102', '054937107', '058498106', '060505104', '064058100', '071813109', '086516101', '097023105', '124857202', '12504L109', '125896100', '127387108', '14149Y108', '143130102', '143658300', '151020104', '15189T107', '156700106', '156782104', '165167107', '166764100', '171340102', '17275R102', '172967424', '177376100', '191216100', '192446102', '194162103', '205887102', '20825C104', '21036P108', '219350105', '22822V101', '23331A109', '235851102', '25179M103', '254687106', '25470M109', '263534109', '26441C204', '26875P101', '26884L109', '285512109', '30161N101', '30212P303', '30219G108', '30231G102', '30303M102', '316773100', '337738108', '337932107', '34959E109', '354613101', '35671D857', '364760108', '369604103', '370334104', '37247D106', '375558103', '382550101', '40412C101', '40434L105', '406216101', '410345102', '42809H107', '436106108', '436440101', '437076102', '438516106', '45337C102', '458140100', '45866F104', '460146103', '46625H100', '478160104', '485865109', '494368103', '49456B101', '500754106', '501044101', '501797104', '518439104', '526057104', '52729N308', '532457108', '548661107', '552953101', '565849106', '56585A102', '571903202', '574599106', '57636Q104', '577081102', '57772K101', '580135101', '582839106', '58933Y105', '594918104', '595017104', '595112103', '61166W101', '61174X109', '617446448', '640268108', '64110L106', '651229106', '651290108', '65339F101', '654106103', '65473P105', '655044105', '674599105', '682680103', '68389X105', '69331C108', '693475105', '693506107', '69351T106', '70450Y103', '713448108', '717081103', '718172109', '718546104', '723787107', '743315103', '744320102', '744573106', '745867101', '747525103', '75281A109', '756577102', '7591EP100', '761713106', '778296103', '79466L302', '806857108', '81211K100', '828806109', '83088M102', '842587107', '871503108', '871829107', '872540109', '87612E106', '87901J105', '880770102', '882508104', '883203101', '88579Y101', '886547108', '887317303', '892356106', '90184L102', '902973304', '911312106', '912909108', '913017109', '91324P102', '92343V104', '92532F100', '925550105', '92826C839', '931142103', '931427108', '949746101', '958102105', '959802109', '962166104', '98212B103', '983134107', '98389B100', '983919101', 'G1151C101', 'G48833100', 'G5960L103', 'G97822103', 'H8817H100', 'N59465109']

2017/06/30: ['001055102', '00130H105', '001547108', '002824100', '00287Y109', '00507V109', '00724F101', '00817Y108', '00846U101', '02209S103', '023135106', '02376R102', '025537101', '025816109', '026874784', '03027X100', '031162100', '037411105', '038222105', '039483102', '052769106', '053774105', '054937107', '058498106', '060505104', '064058100', '071813109', '075887109', '086516101', '093671105', '097023105', '099724106', '101137107', '110122108', '12504L109', '125896100', '126408103', '12673P105', '127097103', '127387108', '14040H105', '143130102', '143658300', '149123101', '15189T107', '156782104', '165167107', '171779309', '17275R102', '172967424', '174610105', '191216100', '192446102', '194162103', '20030N101', '205887102', '20825C104', '219350105', '222070203', '22822V101', '244199105', '247361702', '254687106', '25470F104', '25470M109', '256677105', '25746U109', '263534109', '26441C204', '26875P101', '26884L109', '269246401', '278642103', '285512109', '30161N101', '30212P303', '30219G108', '30303M102', '316773100', '336433107', '337932107', '345370860', '354613101', '36467W109', '37045V100', '37247D106', '382550101', '40412C101', '406216101', '410345102', '416515104', '42824C109', '43300A203', '436440101', '437076102', '438516106', '446150104', '45866F104', '459200101', '460146103', '461202103', '46625H100', '478160104', '48666K109', '493267108', '49456B101', '501797104', '501889208', '512807108', '526057104', '532457108', '552848103', '552953101', '56585A102', '571903202', '574599106', '57636Q104', '57772K101', '580135101', '58933Y105', '59156R108', '595017104', '595112103', '61166W101', '61174X109', '617446448', '626717102', '629377508', '637071101', '63938C108', '64110L106', '651229106', '654106103', '65473P105', '67066G104', '674599105', '676220106', '681919106', '682680103', '68389X105', '69331C108', '693475105', '69351T106', '704326107', '70450Y103', '708160106', '713448108', '717081103', '718172109', '718546104', '74144T108', '743315103', '74340W103', '744320102', '747525103', '79466L302', '806857108', '808513105', '81211K100', '83088M102', '844741108', '855244109', '867914103', '87165B103', '874054109', '88033G407', '90184L102', '902973304', '904311206', '907818108', '910047109', '911312106', '913017109', '91324P102', '918204108', '92532F100', '925550105', '92826C839', '931142103', '949746101', '958102105', '959802109', '962166104', '966837106', '969457100', '983134107', '98389B100', '983919101', '988498101', '98850P109', '989701107', '98978V103', 'G1151C101', 'G30401106', 'G47567105', 'G491BT108', 'G51502105', 'G5960L103', 'G60754101', 'G7665A101']

2017/09/30: ['001055102', '002824100', '00287Y109', '002896207', '00507V109', '00724F101', '00817Y108', '00971T101', '013872106', '01741R102', '021346101', '02209S103', '023135106', '025816109', '031162100', '032511107', '032654105', '037411105', '038222105', '039483102', '03965L100', '052769106', '053774105', '060505104', '071813109', '075896100', '093671105', '097023105', '099724106', '101137107', '110122108', '125269100', '12541W209', '126408103', '126650100', '127097103', '134429109', '143658300', '149123101', '151020104', '156782104', '166764100', '17275R102', '172967424', '174610105', '191216100', '192446102', '20030N101', '20825C104', '219350105', '22160K105', '22822V101', '23331A109', '235851102', '25179M103', '25271C102', '254709108', '25470M109', '256677105', '256746108', '25746U109', '26875P101', '26884L109', '269246401', '278642103', '285512109', '291011104', '294429105', '30219G108', '30303M102', '311900104', '313400301', '316773100', '320517105', '345370860', '364760108', '369604103', '37045V100', '37247D106', '375558103', '38141G104', '40412C101', '40434L105', '406216101', '423452101', '42809H107', '42824C109', '43300A203', '436106108', '437076102', '438516106', '440452100', '446150104', '458140100', '45866F104', '459200101', '460146103', '46625H100', '493267108', '49446R109', '49456B101', '500255104', '500754106', '501044101', '512807108', '518439104', '532457108', '546347105', '548661107', '552848103', '552953101', '55616P104', '565849106', '56585A102', '571903202', '574599106', '57636Q104', '57772K101', '58933Y105', '59156R108', '594918104', '595017104', '595112103', '609207105', '61174X109', '617446448', '61945C103', '626717102', '637071101', '64110D104', '64110L106', '651229106', '651290108', '655044105', '655664100', '67066G104', '67103H107', '674599105', '682680103', '693475105', '704326107', '70450Y103', '712704105', '717081103', '718546104', '723787107', '742718109', '74340W103', '745867101', '747525103', '75281A109', '774341101', '778296103', '79466L302', '806857108', '808513105', '842587107', '845467109', '871503108', '871829107', '872540109', '874054109', '87612E106', '880770102', '882508104', '892356106', '896945201', '90184L102', '902494103', '902973304', '904311206', '907818108', '911312106', '912909108', '91324P102', '917047102', '91913Y100', '92343V104', '92826C839', '931142103', '931427108', '949746101', '962166104', '98212B103', '983134107', '983919101', '98850P109', '989701107', '98978V103', 'G1151C101', 'G29183103', 'G3157S106', 'G48833100', 'G5960L103', 'G6359F103', 'G65431101', 'G7665A101', 'G98294104', 'H8817H100', 'N53745100']

2017/12/31: ['001055102', '00206R102', '002824100', '00287Y109', '002896207', '00507V109', '00724F101', '00817Y108', '00971T101', '021346101', '02209S103', '023135106', '02376R102', '025537101', '025816109', '026874784', '032654105', '037833100', '053774105', '054937107', '058498106', '060505104', '064058100', '071813109', '086516101', '093671105', '097023105', '101137107', '110122108', '125269100', '126408103', '127097103', '134429109', '14040H105', '149123101', '151020104', '15189T107', '156700106', '166764100', '17275R102', '172967424', '174610105', '192446102', '194162103', '20030N101', '205887102', '20825C104', '219350105', '22160K105', '222070203', '23331A109', '235851102', '23918K108', '247361702', '247916208', '25179M103', '254687106', '254709108', '25470F104', '256677105', '256746108', '25746U109', '26441C204', '26875P101', '26884L109', '269246401', '281020107', '291011104', '29414D100', '30212P303', '30219G108', '30303M102', '311900104', '316773100', '320517105', '336433107', '344849104', '345370860', '35671D857', '36174X101', '36467W109', '364760108', '369604103', '370334104', '37045V100', '37247D106', '38141G104', '40412C101', '406216101', '412822108', '436106108', '436440101', '437076102', '438516106', '440452100', '44107P104', '446150104', '458140100', '45866F104', '46625H100', '478160104', '48203R104', '48666K109', '487836108', '493267108', '500255104', '501044101', '501797104', '526057104', '548661107', '552953101', '55616P104', '565849106', '56585A102', '571903202', '574599106', '57636Q104', '580135101', '58155Q103', '58933Y105', '594918104', '595112103', '609207105', '61166W101', '617446448', '61945C103', '626717102', '640268108', '64110D104', '64110L106', '651229106', '651639106', '654106103', '65473P105', '655664100', '670346105', '674599105', '68389X105', '69351T106', '70450Y103', '712704105', '713448108', '717081103', '718172109', '724479100', '742718109', '743315103', '744573106', '745867101', '747525103', '7591EP100', '778296103', '79466L302', '808513105', '81211K100', '83088M102', '842587107', '844741108', '855244109', '867914103', '87165B103', '871829107', '872540109', '87612E106', '87901J105', '882508104', '887317303', '896945201', '90130A101', '90184L102', '902494103', '902973304', '904311206', '907818108', '910047109', '912909108', '913017109', '91324P102', '918204108', '91913Y100', '92343V104', '92826C839', '931142103', '931427108', '949746101', '958102105', '959802109', '962166104', '98212B103', '98978V103', 'G0177J108', 'G491BT108', 'G51502105', 'G5960L103', 'G60754101', 'G6359F103', 'G7945M107', 'G87110105', 'G98294104', 'N59465109']

7.2 List of stocks - Copycat Strategy

2013/12/31: ['478160104', '539830109', '369604103', '437076102', '166764100', '260003108', '30231G102', '718172109', '58933Y105', '054937107', '585055106', '494368103', '88579Y101', '458140100', '110122108', '92343V104', '438516106', '370334104', '20825C104', '191216100', '883556102', '931142103', '713448108', '037411105', '742718109', '459200101', '911312106', '254687106', '863667101', '655664100', '001055102', '594918104', '291011104', '871829107', '00287Y109', '26441C204', '292766102', '872375100', '00206R102', '654902204', '337915102', 'G1151C101', '02209S103', '25746U109', '717081103', '149123101', '609207105', '46625H100', '055622104', '50076Q106', '580135101', '30231G102', '78355W106', '464287655', '464287200', '464287465', '921937827', '46429B663', '464287234', '464287697', '464287168', '73935S105', '68375N103', '464287309', '922042858', '464287408', '92206C664', '464288448', '464289859', '921943858', '40412C101', '464289867', '831865209', '922908413', '78463V107', '97717W406', '471109108', '30231G102', '459200101', 'H84140112', '464288687', '97717W562', '191216100', '842587107', '78464A409', '343498101', '78463X400', '464287739', '580135101', '478160104', '345370860', '931142103', '7591EP100', '493267108', '37247D106', '465790301', '713448108', '037411105', '594918104', '25179M103', '585055106', '717081103', '58933Y105', '189054109', '931142103', '478160104', '458140100', '87612E106', '02209S103', '577081102', '30231G102', '718172109', '609207105', '655844108', '073730103', '574599106']

2014/03/31: ['12008R107', '698477106', '527288104', '502160104', '050095108', '149568107', '682129101', '007973100', '136385101', '830566105', '423452101', '533900106', '128030202', '374689107', '903002103', '811904101', '032159105', 'N93540107', '817070501', '384313102', '125919308', '481165108', '693366205', 'Y8162K121', '26168L205', '741511109', 'Y81669106', '532746104', '26885G109', '98212B103', '190345108', '60740F105', '12531Q105', '960413102', '37941P306', '45172L209', '055639108', '69336T106', '86101X104', '057755100', '40609P105', '830830105', '874083108', '87425E103', '638522102', '09624H109', '037833100', '887389104', '26852W103', '85814R107', '30231G102', '872540109', '57636Q104', '278865100', '913017109', '037833100', '171340102', '654106103', '151020104', '194162103', '191216100', '580135101', '461202103', '459200101', '053015103', '771195104', '858912108', '826552101', '74005P104', '579780206', '742718109', '303075105', '931422109', '65339F101', '115637209', 'G1151C101', '747525103', '337738108', '26875P101', '244199105', '713448108', '73640Q105', '384802104', '863667101', '594918104', '166764100', '896239100', '670100205', '00206R102', '871829107', '478160104', '462726100', '343498101', '20825C104', '002824100', '249030107', '023135106', '69840W108', '45168D104', '369604103', '30219G108', '883556102', '375558103', '346375108', '92552V100', '317485100', '740189105', '693282105', '097023105', '86800U104', '286082102', '42809H107', '292475209', '032744104', '404052102', '404053100', '285229100', '466313103', '00971T101', '29362U104', '143130102', '20030N101', '438516106', '902104108', '55306N104', '651824104', 'M7516K103', '075887109', '92220P105', '74874Q100', '61166W101', '207410101', '989855101', 'G2554F113', '674001201', '068323104', '50540R409', '114340102', '10921T101', '910047109', '29384R105', '98212B103', '930427109', '482738101', '402635304', '037833100', '87305R109', '101388106', '75936P105', '253393102', '92204A702', '922908751', '478160104', '674599105', '922908629', '031100100', '913017109', '235851102', '037833100', '883556102', '918204108', '747525103', '806857108', '855244109', '78462F103', '609207105', '654106103', '018490102', '166764100', '78464A698', '03027X100', '037411105', '98978V103', '345370860', '33734X192', '369604103', '713448108', '922908637', '594918104', '427866108', '254687106', '87612E106', '922908769', '717081103', '92343V104', '31428X106', '459200101', '863667101', '00724F101', '742718109', '744320102', '458140100', '941848103', '464287549', '580135101', '194162103', '718172109', '278642103', '922908553', '228368106']

2014/06/30: ['115637209', '478160104', '464287473', '464287598', '464287648', '742718109', '464287614', '464287630', '464287465', '464287481', '037833100', '459200101', '369604103', '713448108', '931422109', '166764100', '053015103', '018490102', '872540109', '464287499', '194162103', '580135101', '594918104', '464287655', '458140100', '717081103', '002824100', '478366107', '755111507', '254687106', '464287234', '693475105', '698813102', '674599105', '171232101', '191216100', '718172109', '902973304', '655844108', '444859102', '693506107', '110122108', '904784709', '151020104', '437076102', '009158106', '907818108', '718546104', '949746101', '941848103', '494550106', '260003108', '452308109', '806857108', '681919106', '172062101', '931142103', '871829107', '149123101', '780259206', '487836108', '345370860', '038222105', '235851102', '126650100', '335720108', '464287614', '464287598', '464287507', '464287804', '037833100', '907818108', '57636Q104', '747525103', '38259P508', '949746101', '166764100', '14040H105', '580135101', '464287465', '254687106', '464287234', '151020104', '713448108', '09247X101', '718172109', '741503403', '167250109', '806857108', '478160104', '369604103', '149123101', '02209S103', '278642103', '776696106', '46625H100', '92826C839', '87612E106', '032511107', '30231G102', '191216100', '742718109', '913017109', '464286665', '931142103', '212015101', '74144T108', '92343V104', '23918K108', '88579Y101', '911312106', '855244109', '278865100', '375558103', '459200101', '464287499', '037411105', '902973304', '594918104', '00287Y109', '291011104', '192446102', '09062X103', '437076102', '023135106', '517834107', '464287390', '118230101', '194162103', '20825C104', '002824100', '031162100', '263534109', '61166W101', '337738108', '494550106', '097023105', '134429109', '17275R102', '29250R106', '31428X106', '559080106', '293792107', '726503105', '891160509', '886547108', '524660107', '920253101', '854502101', '717081103', '464287879', '071813109', '110122108', '00206R102', '018490102', '857477103', '055622104', '46428R107', '896522109', '268648102', '534187109', '126650100', '22160K105', '585055106', '842587107']

2014/09/30: ['78464A763', '756207106', '73935X583', '922042874', '00162Q866', '233051200', '922042858', '464286848', '81369Y803', '464286665', '464286509', '81369Y209', '81369Y407', '81369Y605', '81369Y506', '73935X567', '464287861', '81369Y704', '922908629', '81369Y308', '97717W851', '464287234', '922042775', '10807M105', '464287788', '40637H109', '78464A391', '46429B366', '78468R408', '464287622', '464288182', '09062X103', '78467V103', '81369Y100', '73935S105', '81369Y886', '72201R718', '61166W101', '817337405', '72201R775', '922908637', '21925Y103', '75915K101', '464288570', '871639308', '015271109', '83415T101', '037833100', '452536105', '55264U108', '64125C109', '040047102', '44984A105', '97717W307', '452327109', '922908553', '88160R101', '72201R833', '747525103', '922042742', '89484Q100', '192108504', '249908104', '03832V109', '496902404', '422704106', '758338107', '928972108', '8342EP107', '01535P106', '29843R100', '69924P102', '335720108', '464287614', '464287598', '464287507', '037833100', '464287804', '907818108', '57636Q104', '166764100', '949746101', '151020104', '747525103', '14040H105', '580135101', '464287465', '254687106', '713448108', '464287234', '09247X101', '718172109', '741503403', '806857108', '478160104', '38259P508', '369604103', '02209S103', '149123101', '776696106', '167250109', '92826C839', '032511107', '46625H100', '278642103', '30231G102', '742718109', '212015101', '913017109', '87612E106', '191216100', '375558103', '88579Y101', '74144T108', '931142103', '23918K108', '92343V104', '037411105', '464286665', '517834107', '855244109', '902973304', '911312106', '278865100', '459200101', '464287499', '594918104', '023135106', '00287Y109', '437076102', '09062X103', '291011104', '20825C104', '118230101', '192446102', '194162103', '031162100', '464287390', '002824100', '494550106', '61166W101', '31428X106', '263534109', '29250R106', '337738108', '097023105', '17275R102', '559080106', '293792107', '134429109', '018490102', '886547108', '891160509', '726503105', '854502101', '00206R102', '524660107', '717081103', '464287879', '920253101', '071813109', '110122108', '055622104', '896522109', '857477103', '22160K105', '46428R107', '534187109', '585055106', '165167107', '74005P104', '941848103']

2014/12/31: ['750469207', '29428V104', '737010108', '89685A102', '03874P101', '71714G102', '15117E107', '20600T108', '92337C104', '576314108', '29365M208', '800677106', '556099109', '949746101', '902973304', '025816109', '46625H100', '38141G104', '09247X101', '857477103', '907818108', '693475105', 'G0408V102', '615369105', '87236Y108', '14040H105', '06652K103', '57636Q104', '743606105', '55261F104', '054937107', '38259P508', '436893200', '861025104', '064058100', 'G32335104']

2015/03/31: ['758932107', '93317W102', '81616X103', '09247X101', '16208T102', '016230104', '29084Q100', '85571B105', 'G81276100', '26969P108', '29362U104', '125581801', '855030102', '76131N101', '85571W109', '74838C106', '663904209', 'N71546100', '699374302', '82621J105', '737010108', '868459108', 'G55598109', '74371L109', '253031108', '15130J109', '023111206', '75915K101', '98978L105', '44915N101', '68230A106', '03674U201']

2015/06/30: ['550021109', '398438408', '806857108', 'N07059210', 'N22717107', '874039100', '670100205', '01609W102', '22943F100', 'G27823106', 'Y0486S104', '75886F107', '57636Q104', '741503403', '74144T108', '151020104', '04685W103', '747525103', '023135106', '38259P508', '037833100', '49271M100', '98138H101']

2015/09/30: ['501889208', '931427108', '267475101', '31428X106', '00751Y106', '00101J106', '60740F105', '695156109', '184496107', '59522J103', '950755108', '885160101', '87074U101', '573075108', '15201P109', '115637209', '698813102', '30231G102', '478160104', '037833100', '922908744', '78464A607', '742718109', '464287309', '713448108', '369604103', '931427108', '922908512', '464287598', '33734X200', '464287614', '464287481', '464287606', '464287473', '464287648', '00287Y109', '464287630', '459200101', '872540109', '053015103', '191216100', '46625H100', '194162103', '166764100', '88579Y101', '464287465', '458140100', '254687106', '68389X105', '717081103', '78464A300', '002824100', '580135101', '464287176', '594918104', '444859102', '755111507', '478366107', 'G0177J108', '693475105', '78463V107', '437076102', '26441C204', '151020104', '78464A201', '58933Y105', '110122108', '693506107', '375558103', '20030N101', '902973304', '806857108', '345370860', '20825C104', '26884L109', '718172109', '22160K105', '00206R102', '02376R102', '941848103', '78468R408', '78462F103', '45071R109', '037833100', '855244109', '73935A104', '464287655', '58933Y105', '713448108', '97717W851', '742718109', '78464A763', '872540109', '74005P104', '81369Y803', '38259P508', '026874784', '126650100', '883556102', '949746101', '172967424', '858912108', '902973304', '674599105', '22160K105', '192446102', '031100100', '464288513', '002824100', '235851102', '459506101', '406216101', '92927K102', '81369Y506', '88160R101', '26875P101', '30303M102', '81369Y605', '032511107', '741503403', '464286806', '464287291', '055622104', '81369Y209', '110122108', '464286772', '81369Y407', '191216100', '464288695', '023135106', '873868103', '747525103', '911163103', '775043102', '18383Q135', '254709108', '30219G108', '885160101', '84610H108', '97717X701', '464287168', '464287556', '500634209', '966837106', '464286103', '464288174', '464286749']

2015/12/31: ['697660207', 'L00849106', '879273209', '893870204', '71646J109', '984245100', '399909100', '97717W851', '46434V704', '05961W105', '450047204', '27888D101', '07329M100', '891027104', '29244A102', '880890108', 'G38327105', '037833100', '97717W422', 'G0457F107', '697900108', '78462F103', '464286400', '463588103', '655844108', '88160R101', '742718109', '437076102', '30231G102', '254687106', '59522J103', '921908844', '001055102', '478160104', '73935X716', '670346105', '00287Y109', '458140100', '931427108', '539830109', 'G0177J108', '191216100', '14149Y108', '02209S103', '580135101', '91324P102', '842587107', '88579Y101', '112585104', '37733W105', '858912108', '78463X202', '291011104', '74347R248', '755111507', '370334104', '94106L109', '718172109', '654106103', '03524A108', '235851102', '806857108', '57060U605', '7591EP100', '74874Q100', '378973408', '30226D106', '747906204', '31620M106', '871829107', '478160104', '718172109', '369604103', '713448108', '88579Y101', '372460105', '460146103', '00287Y109', '30231G102', '166764100', '423452101', '742718109', '87233Q108', '539830109', '20825C104', '87611X105', '717081103', '370334104', '580135101']

2016/03/31: ['594918104', '311900104', '056752108', '02079K107', '225310101', '67011P100', '244199105', '898202106', '25659T107', '570535104', '527288104', '637071101', '235851102', '023135106', '617700109', '92206C409', '78467Y107', '464288646', '464288158', '464288414', '92206C870', '922908595', '037833100', '05617K109', '464287176', '30231G102', '464287465', '922908363', '808524508', '808524300', '729251108', '594918104', '458140100', '449172105', '478160104', '023135106', '369604103', '92343V104', '78464A425', '713448108', '808524102', '872540109', '02079K107', '00206R102', '808524805', '375558103', '254687106', '808524607', '78355W593', '126650100', '440452100', '40414L109', '46625H100', '17275R102', '949746101', '808524839', '913017109', '742718109', '464288687', '464287572', '063671101', '717081103', '806857108', '052769106', '902973304', '88579Y101', '437076102', '636274300', '46429B655', '808524847', '832696405', '629579103', '464286509', '892356106', '067383109', '871829107', '194162103', '464289529', '48248M102', '78355W106', '166764100', '00817Y108', '30303M102', '002824100', '263534109', '260543103', '532457108', '835495102', '500643200', '05616B100', '00287Y109', '83175M205', '857477103', '477143101', '45780L104', '92345Y106', '150870103', '896522109', '268648102', '790849103', '580135101', '65339F101', 'G54050102', '57060U803', '855244109', '370334104', '87612E106', '452308109', '110122108', '73935X575', '438516106', '46131B100', '20030N101', '808524201', '907818108']

2016/06/30: ['92339V100', '02005N100', '125269100', 'G00349103', '235825205', '026874784', 'Y81669106', '02376R102', '247361702', '550678106', '910047109', '45031U101', 'G27823106', '629209305', '844741108', '37247D106', '984245100', '896047503', '966387102', '78463V107', '011532108', '498696103', '04341Y105', '780287108', '82823L106', '008474108', '284902103', '828336107', '651639106', '035128206', '57060U100', '927926303', '697900108', '301835104', '205750102', '74164W106', '067901108', '88741P103', 'G3788M114', '16411W108', '900435108', 'N00985106', '595112103', '16411R208', '969457100', 'Y2106R110', '026874784', '527288104', 'G3922B107', '43289P106', '06740L519', '896047503', '45104G104', '617446448', '876108101', '76117W109', 'M6786D104', '358010106', '41902R103', '86722A103', '278768106', '46131B100', '74972G103', '046220109', '02503Y103', '04878Q863', '10807M105', '294628102', '141665109', '960878106', '482539103', '03763V102', '78377T107', '12811R104', '31679F101', '31678A103', '85814R107', '69327R101', '164651101', '651290108', '037347101', '87244T109', 'G67506108', '58503T106', '50077C106', '58503F106', '31679P109', '629156407', '12768T103', '68620A104']

2016/09/30: ['04351G101', '12008R107', '08915P101', '30219E103', '118440106', '760281204', '727493108', '422819102', '913915104', '320218100', '94188P101', '83013P105']

2016/12/31: ['457187102', '194162103', '440452100', '053015103', '337738108', '681919106', '370334104', '609720107', '118230101', '87233Q108', '400506101', '579780206', '030975106', '038336103', '081437105', '743606105', '891160509', '293792107', '704326107', '74005P104', '655844108', '291011104', '384109104', '009158106', '384802104', '86764L108', '580135101', '592688105', '701094104', '494368103', '68268N103', '98419M100', '90337L108', '067383109', '459506101', '636274300', '45073V108', '824348106', '053611109', '376536108', '66987V109', '726503105', 'G51502105', '756109104', '032511107', '438516106', '524660107', '23311P100', '917286205', '165303108', '339041105', '34984V100', '02079K305', '64110L106', '741503403', '47215P106', '22943F100', 'G3922B107', '74975N105', '29786A106', '92826C839', '57777M201', '530307305', 'G47567105', '294628102', '45672H104', '339041105', '741503403', '92826C839', '65336K103', 'G96629103', '09238E104', '058498106', '62944T105', '278715206', '59100U108', '04010E109', '925815102', '54142L109', '217204106', '53224V100', '868459108', '584688105', '92552V100', '57060D108', '58471A105', '78486Q101', '45337C102', '639050103', '53220K504', '163092109', '64829B100', '902252105', '922908769', '537008104', '75954W107', 'G47567105', '29404K106', '22160N109', '464287226', '317485100', '87651B104', '351858105', '23335C101', '62945V109', '73935A104', '907818108', '562750109', '501889208', '852857200', '023586100', '11273Q109', '464287242', '057665200', '30231G102', '233912104', 'G35569105', '776696106', '00163U106', '464287465', '464288620', '56165R103', '73936T474', '464288885']

2017/03/31: ['13321L108', '828336107', '496902404', '780287108', '067901108', '900435108', '698354107', 'N97284108', 'Y2066G104', 'G9108L108', '48268K101', '66510M204', '48242W106', '380956409', '18911Q102', '66987E206', '644535106', '29082A107', 'Y7388L103', 'G63907102', '15234Q207', '852234103', '03783C100', '47215P106', '29978A104', '31680Q104', '09239B109', '01609W102', '81762P102', '50212V100', '011642105', '64110L106', '30303M102', '530307305', 'G06242104', '88339J105', 'V5633W109', '22266L106', '04650Y100', '26856L103', '28849P100', '110394103', '34354P105', '594918104', '68389X105', '718172109', '075887109', '883556102', '194162103', '02209S103', '00206R102', '369604103', '717081103', '650111107', '055622104', '436440101', '55262C100', '649445103', '811543107', '37733W105', '09228F103', '717081103', '703481101', '85917A100', 'G63907102', '89656D101', '91829F104', '862578101', '44951N106', '63938C108', '66987V109', '110122108', '30231G102', '20030N101', '459200101', '680223104', '060505104', '166764100', '98923T104', '369604103', '046265104']

2017/06/30: ['30303M102', 'N20146101', '256677105', '249906108', '112585104', '776696106', '03662Q105', '98139A105']

2017/09/30: ['Y7546A122', '294100102', '55933J203', 'G3682E192', '50077C106', '023139884', 'N22035104', '69888T207', 'Y7388L103', 'Y62267102', '88160R101', '01609W102', '874080104', '47759T100', '163075104', '023135106', '30303M102', '33733E203', '75886F107', '60879B107', '22943F100', '47215P106', '00461U105', '84652J103', 'H5861P103', '92529L102', '632307104', 'H03579101', '60470J103', 'G4783X105', '43358V109', '026874784', '552848103', 'G0585R106', '00912X302', '85571B105', '530307305', '89679E300', '762760106', '237266101', '053332102', '85711T305', '693366205', '14040H105', '01861G100', '476405105', '862121100', '680223104', '67091N108', 'G3223R108', 'Y62132108', 'G7496G103', '860372101', 'G67742109', 'G9618E107', '191216100', '320517105', '035710409', '037833100', '7591EP100', '478160104', '718172109', '594918104', '58933Y105', '985817105', '697435105', '64110L106', '056752108', '052769106', '192446102', '70450Y103', '00971T101', '22266L106', '37518B102', '78410G104', '21925Y103', '594972408', '78462F103', '67066G104', '012653101', '023135106', '833635105', '538034109', '848637104', 'G96629103', '03027X100', '053015103', '609207105', '931427108', '501044101', '478160104', '78410G104', '832696405', '110122108', '189054109', '12508E101']

2017/12/31: ['032724106', '03152W109', '451731103', '02083G100', '19459J104', '462260100', '82640U107', '171757107', '05591B109', '12008R107', '960413102', '149568107', '502160104', '698477106', '88642R109', '65548P403', '527288104', '050095108', '30227H106', '682129101', '750491102', '591520200', '533900106', '903002103', '136385101', '45774W108', '05491N104', '958102105', '032159105', '811904101', '423452101', '693366205', '374689107', '235825205', 'Y81669106', 'Y2106R110', '02553E106', '98416J118', '830830105', '78413P101', '239360100', '98212B103', '741511109', '007973100', '638517102', '887389104', '053807103', '231021106', '101044105', '949746101', '03062T105', '69007J106', '903293405', '65373J209', '01877R108', '64031N108', '01861G100', '021346101', '02005N100', '984245100', '16411R208', '29250R106', '450047204', 'G81477104', '698477106', '458118106', '61022P100', '868873100', '307675108', '482564101', '87874R100', '98421B100', '37253A103', '82817Q103', '125141101', '404609109', '746228303', '607525102', '25659T107', '06647F102', '13100M509']

References

[1] Ben-David, Itzhak, Denys Glushkov, and Rabih Moussawi: Do arbitrageurs really avoid high idiosyncratic risk stocks?

[2] Cohen, Folk, Silli: Best Ideas

[3] Myers, Poterba, Shackeford, Shoven: Copycat Funds: Information Disclosure Regulation and the Returns to Active Management in the Mutual Fund Industry

[4] Wermers: Is Money Really 'Smart'? New Evidence on the Relation Between Mutual Fund Flows, Manager Behavior, and Performance Persistence

[5] Verbeek, Wang: Better than the Original? The Relative Success of Copycat Funds

[6] Goldman Sachs Quantitative Investment Strategies: What does the X-Ray Machine Reveal About Hedge Funds

[7] Sirri, Tufano: Costly Search and Mutual Fund Flows

[8] Goldman Sachs Quantitative Investment Strategies: What does the X-Ray Machine Reveal About Hedge Funds

Acknowledgement

At the end, we want to express our grateful to all the supports we received from our sponsors and advisor. Thank you Mr. Alexander Fleiss for sponsoring our project and providing keen market insights that helped us analyze data and build various features. Thank you Mr. Han Cui for providing inspiring ideas and technical support. Thank you Professor Sasha Stoikov for putting forward valuable suggestions and assisting us in organizing the research process.